MARYGROVE COLLEGE
LIBRARY
DETROIT, MICH.

Child Development

me 79 Number 4

August 2008

Editor

Lynn S. Liben

Associate Editors

Joseph P. Allen
Neil E. Berthier
Marvin W. Daehler
Nancy E. Hill
Kathy A. Hirsh-Pasek
Campbell Leaper
Patricia H. Miller
Sharon Nelson-Le Gall
Paul C. Quinn
Theresa A. Thorkildsen

CURRENT YR/VOL

Marygrove College Library
8425 West McNichols Road
Detroit, MI 48221

Published by Blackwell Publishing for the Society for Research in Child Development

Published by Blackwell Publishing for the Society for Research in Child Development

Child Development

Child Development (ISSN 0009-3920), one of three publications of the Society for Research in Child Development, Inc., is published bimonthly by Blackwell Publishing, Inc., with offices at 350 Main Street, Malden, MA 02148, USA; 9600 Garsington Road, Oxford OX4 2ZG, UK; and 155 Cremorne St., Richmond VIC 3121, Australia. Blackwell Publishing, Inc., is now part of Wiley-Blackwell. Institutional Premium Subscription Rate (includes online access to current content and all online back files to January 1st 1997, where available): The Americas, $585; Rest of the World, £417. Premium Online-Only rates are also available (see below). Prices include delivery of print journals to the recipient's address. Delivery terms are Delivered Duty Unpaid (DDU); the recipient is responsible for paying any import duty or taxes. Legal title passes to the customer on despatch by our distributors.

Customers in Canada should add 5% GST to the America rate or provide evidence of entitlement to exemption.

Customers in the UK and EU should add VAT at 7% to the rest of the World rate or provide a VAT registration number or evidence of entitlement to exemption.

For more information about Blackwell Publishing journals, including online access information, terms and conditions, and other pricing options, please visit www.blackwellpublishing.com or contact our customer service department, tel: (800) 835-6770 or + (781) 388-8599 (US office) +44 (0)1 865 778315 (UK office) or +65 6511 8000 (Asia Office). Subscription in 2008 includes three issues of *Monographs of the Society for Research in Child Development* and three issues of *Child Development Perspectives*. Single issues from current and recent volumes are available at the current single issue price from Blackwell Publishing Journals. Earlier issues may be obtained from Periodicals Service Company, 11 Main Street, Germantown, NY 12526, USA. Tel: +1 518 537 4700, Fax: +1 518 537 5899, Email: psc@periodicals.com. Volumes 37 to date are available in microfilm from ProQuest Information and Learning, Serials Customer Service Department, 300 North Zeeb Road, Ann Arbor, MI 48106; telephone: (800) 521-0600 x2873.

Electronic Access: The *Child Development* website can be found at www.srcd.org/cd.html. Abstract information for this journal is electronically available at www.blackwellpublishing.com/cdev. For information on full-text access, see www. blackwellpublishing.com.

Postmaster: Periodical postage paid at Boston, MA and additional offices. Mailing to the rest of the world by International Mail Express (IMEX). Canadian mail is sent by Canadian Publications Mail agreement number 40064709. Postmaster: Send all address changes to *Child Development*, Blackwell Publishing Inc., Journals Subscription Department, 350 Main St., Malden, MA 02148-5020.

For new orders, renewals, sample copy requests, claims, changes of address and all other subscription correspondences please contact the Journals Department at your nearest Blackwell office (address details listed above). UK office phone: +44 (0) 1865-778315, Fax: +44 (0) 1865-471775, Email: customerservices@blackwellpublishing.com; US office phone: 800-835-6770 or 781-388-8599, Fax: 781-388-8232, Email: customerservices@blackwellpublishing.com; Asia office phone: +65 6511 8000, Fax: +44 (0) 1865-471775, Email: customerservices@blackwellpublishing.com

Membership communication and address changes should be addressed to the SRCD Executive Office, 2950 South State Street, Suite 401, Ann Arbor, MI 48104, USA. http://www.srcd.org/

Journal Compilation © 2008 The Society for Research in Child Development, Inc. All rights reserved. With the exception of fair dealing for the purposes of research or private study, or criticism or review, no part of this publication may be reproduced, stored or transmitted in any form or by any means without the prior permission in writing from the copyright holder. Authorization to photocopy items for internal and personal use is granted by the copyright holder for libraries and other users of the Copyright Clearance Center (CCC), 222 Rosewood Drive, Danvers, MA 01923 (www.copyright.com), provided the appropriate fee is paid directly to the CCC. This consent does not extend to other kinds of copying, such as copying for general distribution for advertising or promotional purposes, for creating new collective works or for resale. Institutions with a paid subscription to this journal may make photocopies for teaching purposes free of charge provided such copies are not resold. Special request should be addressed to journalsrights@oxon.blackwellpublishing.com.

For advertising information, please visit the journal's website at www.blackwellpublishing.com/cdev or contact the Academic and Science Advertising Sales Coordinator at journaladsUSA@bos.blackwellpublishing.com. 350 Main St., Malden, MA 02148. Phone: 781.388.8532, Fax: 781.338.8532.

Sign up to receive Blackwell *Synergy* free e-mail alerts with complete *Child Development* tables of contents and quick links to article abstracts from the most current issue. Simply go to www.blackwell-synergy.com, select the journal from the list of journals, and click on "Sign-up" for FREE email table of contents alerts.

Indexing/Abstracting: The contents of this journal are indexed or abstracted in the following: *Academic ASAP; Expanded Academic ASAP; Academic Search Elite; Academic Search Premier; Applied Social Sciences Index and Abstracts; Biomedical Reference Collection: Comprehensive Edition; CAB Abstracts; CABDirect; Contents Pages in Education; Current Contents/Social & Behavioral Science; Current Index to Journals in Education; Cambridge Scientific Abstracts Linguistics & Language Behavior Abstracts; CatchWord; Current Index to Journals in Education; Educational Research Abstracts; Environmental Sciences & Pollution Management; e-psyche; EBSCO Sociological Collection; EBSCO Online; EBSCO Professional Development Collection; Educational Research Abstracts; Educational Resources Information Center; Health and Safety Science Abstracts; ISI Basic Social Sciences Index; InfoTrac OneFile; Ingenta; JCR Social Sciences Edition; Leisure, Recreation, and Tourism Abstracts; Online Computer Library Center FirstSearch Electronic Collections Online; Periodical Abstracts Library; Periodical Abstracts Research II; ProQuest 5000; ProQuest Social Science Plus Text; PsycINFO; Psychological Abstracts; Psychology and Behavioral Sciences Collection; Pollution Abstracts; Safety Science & Risk Abstracts; Sage Family Studies Abstracts; Social Services Abstracts; Social Science Citation Index; Sociological Abstracts; Student Resource Center Bronze; Student Resource Center Gold; Student Resource Center Silver; Violence & Abuse Abstracts; Wilson OmniFile V; Wilson Social Sciences Abstracts;* and *Wilson Social Sciences Index.*

The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials, ANSI Z39.48—1984. This journal is printed on acid-free paper.

Published by Blackwell Publishing for the Society for Research in Child Development

Child Development

Editor

Lynn S. Liben
The Pennsylvania State University

Associate Editors

Joseph P. Allen, *University of Virginia*
Neil E. Berthier, *University of Massachusetts*
Marvin W. Daehler, *University of Massachusetts*
Nancy E. Hill, *Duke University*
Kathy A. Hirsh-Pasek, *Temple University*
Campbell Leaper, *University of California, Santa Cruz*
Patricia H. Miller, *The University of Georgia*
Sharon Nelson-Le Gall, *University of Pittsburgh*
Paul C. Quinn, *University of Delaware*
Theresa A. Thorkildsen, *University of Illinois at Chicago*

Managing Editor

Detra N. Davis

Editorial Assistant

Kaitlin MacKenzie

Consulting Editors

Brian P. Ackerman
Daisuke Akiba
Eric M. Anderman
Roger Bakeman
Lynne E. Baker-Ward
Karen Bartsch
Amy D. Bellmore
Ramesh S. Bhatt
Rebecca S. Bigler
Pamela Blewitt
Celia A. Brownell
Peter Bryant
Kristin Buss
Susan D. Calkins
Tara C. Callaghan
Joseph J. Campos
Deborah M. Capaldi
Noel A. Card
Zhe Chen
Pamela M. Cole
John Colombo
William E. Cross Jr.
Pamela E. Davis-Kean
Adele Diamond
Mary Ann Foley
Bridget Franks
Andrew J. Fuligni
Mary Gauvain
Susan L. Golbeck
H. Hill Goldsmith
Roberta M. Golinkoff
Douglas Granger
Catherine A. Haden
Laura D. Hanish
Charles C. Helwig
Gail D. Heyman
Erika Hoff
George J. Hollich
Alice Sterling Honig
Cynthia Hudley
Jennifer Jenkins
Susan Scanlon Jones
Deborah G. Kemler Nelson
Melanie Killen
Roger Kobak
Grazyna Kochanska
Brett Laursen
Bonnie Jean Leadbeater
Elizabeth Ann Lemerise
Yonata Levy
Joan M. Lucariello
Sarah C. Mangelsdorf
Denis Mareschal
Herb W. Marsh
Clay W. Mash
Susan M. McHale
Derek E. Montgomery
Ulrich Mueller
Tamera Murdock
Letitia R. Naigles
Ageliki Nicolopoulou
Thomas G. O'Connor
Yukari Okamoto
Jamie M. Ostrov
David G. Perry
Jodie M. Plumert
Gary Resnick
Lori A. Roggman
Glenn I. Roisman
Amanda J. Rose
Kenneth H. Rubin
Diane N. Ruble
Martin D. Ruck
Ted Ruffman
Dale H. Schunk
Margaret L. Signorella
Judith G. Smetana
David Sobel
Christia Spears Brown
Cynthia A. Stifter
Mike L. Stoolmiller
Susan Sy
Lisa E. Szechter
Rebecca Treiman
Georgene L. Troseth
Marion K. Underwood
Marinus H. van IJzendoorn
Brian E. Vaughn
Melissa K. Welch-Ross
Allan Wigfield
Terry Wood
Jacqueline D. Woolley
Barbara Younger-Rossmann

Child Development, July/August 2008, Volume 79, Number 4, Pages 807–1184

Contents

EMPIRICAL ARTICLES

Congruency as a Nonspecific Perceptual Property Contributing to Newborns' Face Preference
Viola Macchi Cassia, Eloisa Valenza, Francesca Simion, and Irene Leo 807

Grownups Are Not Afraid of Scary Stuff, but Kids Are: Young Children's and Adults' Reasoning About Children's, Infants', and Adults' Fears
Liat Sayfan and Kristin Hansen Lagattuta 821

Gender and Patterns of Concerned Responsiveness in Representations of the Mother–Daughter and Mother–Son Relationship
Ruth Butler and Rachel Shalit-Naggar 836

Development of Feedforward Control in a Dynamic Manual Tracking Task
Dominique van Roon, Karen Caeyenberghs, Stephan P. Swinnen, and Bouwien C. M. Smits-Engelsman 852

Social and Cognitive Correlates of Children's Lying Behavior
Victoria Talwar and Kang Lee 866

Moral Development in a Violent Society: Colombian Children's Judgments in the Context of Survival and Revenge
Roberto Posada and Cecilia Wainryb 882

Pathways to Literacy: A Study of Invented Spelling and Its Role in Learning to Read
Gene Ouellette and Monique Sénéchal 899

Coaching, Truth Induction, and Young Maltreated Children's False Allegations and False Denials
Thomas D. Lyon, Lindsay C. Malloy, Jodi A. Quas, and Victoria A. Talwar 914

Genetic Influences on Activity Level in Early Childhood: Do Situations Matter?
Kimberly J. Saudino and Jeffrey A. Zapfe 930

Stability and Change in Ethnic Labeling Among Adolescents From Asian and Latin American Immigrant Families
Andrew J. Fuligni, Lisa Kiang, Melissa R. Witkow, and Oscar Baldelomar 944

A Developmental Perspective on Full- Versus Part-Day Kindergarten and Children's Academic Trajectories Through Fifth Grade
Elizabeth Votruba-Drzal, Christine P. Li-Grining, and Carolina Maldonado-Carreño 957

Novel Noun and Verb Learning in Chinese-, English-, and Japanese-Speaking Children
Mutsumi Imai, Lianjing Li, Etsuko Haryu, Hiroyuki Okada, Kathy Hirsh-Pasek, Roberta Michnick Golinkoff, and Jun Shigematsu 979

Does Chronic Classroom Peer Rejection Predict the Development of Children's Classroom Participation During the Grade School Years?
Gary W. Ladd, Sarah L. Herald-Brown, and Mark Reiser 1001

Numerical Magnitude Representations Influence Arithmetic Learning
Julie L. Booth and Robert S. Siegler 1016

Investigating the Development of Data Evaluation: The Role of Data Characteristics
Amy M. Masnick and Bradley J. Morris 1032

Intergenerational Transmission of Social Anxiety: The Role of Social Referencing Processes in Infancy
Lynne Murray, Marc de Rosnay, Joanna Pearson, Caroline Bergeron, Elizabeth Schofield, Melanie Royal-Lawson, and Peter J. Cooper 1049

Family Resources and Parenting Quality: Links to Children's Cognitive Development Across the First 3 Years
Julieta Lugo-Gil and Catherine S. Tamis-LeMonda 1065

Phonological and Semantic Priming in Children With Reading Disability
Rebecca S. Betjemann and Janice M. Keenan 1086

Individual Differences in Adolescents' Beliefs About the Legitimacy of Parental Authority and Their Own Obligation to Obey: A Longitudinal Investigation
Nancy Darling, Patricio Cumsille, and M. Loreto Martínez 1103

The Breadth of Peer Relationships Among Preschoolers: An Application of the Q-Connectivity Method to Externalizing Behavior
Laura D. Hanish, Carol Lynn Martin, Richard A. Fabes, and Hélène Barcelo 1119

The Effects of Background Television on the Toy Play Behavior of Very Young Children
Marie Evans Schmidt, Tiffany A. Pempek, Heather L. Kirkorian, Anne Frankenfield Lund, and Daniel R. Anderson 1137

German Children's Comprehension of Word Order and Case Marking in Causative Sentences
Miriam Dittmar, Kirsten Abbot-Smith, Elena Lieven, and Michael Tomasello 1152

The Dynamics of Life Stressors and Depressive Symptoms in Early Adolescence: A Test of Six Theoretical Models
Margaret Clements, J. Lawrence Aber, and Edward Seidman 1168

OTHER

Manuscripts Accepted 1183

Child Development, July/August 2008, Volume 79, Number 4, Pages 807–820

Congruency as a Nonspecific Perceptual Property Contributing to Newborns' Face Preference

Viola Macchi Cassia
University of Milano-Bicocca

Eloisa Valenza, Francesca Simion, and Irene Leo
University of Padova

Past research has shown that top-heaviness is a perceptual property that plays a crucial role in triggering newborns' preference toward faces. The present study examined the contribution of a second configural property, *congruency,* to newborns' face preference. Experiments 1 and 2 demonstrated that when embedded in nonfacelike stimuli, congruency induces a preference of the same strength as that induced by facedness. Experiments 3 and 4 demonstrated that the attentional biases toward facedness *and* congruency produce a cumulative effect on newborns' visual preferences according to an additive model. These findings were extended by those of Experiment 5, showing that the additive model holds true when congruency is added to top-heaviness in nonfacelike stimuli displaying more elements in the upper portion.

Face recognition in human adults is considered to be a sophisticated and specialized ability. This is because faces are highly complex visual stimuli whose processing involves specific brain regions (Kanwisher, McDermott, & Chun, 1997; Sergent, Ohta, & MacDonald, 1992) and specialized perceptual strategies (see Peterson & Rhodes, 2003, for a review). This specialized response toward faces has sparked debates as to whether specialization may be innate and specific to faces (e.g., Farah, Rabinowitz, Quinn, & Liu, 2000) or may be acquired through the extensive experience individuals acquire with faces within the human species-specific environment (expertise hypothesis; Diamond & Carey, 1986; Gauthier & Tarr, 1997, 2002). Central to this debate is evidence provided by developmental literature showing that just a few days after birth, when visual experience with faces is still minimal, newborns' visual attention is preferentially attracted toward both schematic facelike configurations (Johnson & Morton, 1991; Mondloch et al., 1999; Valenza, Simion, Macchi Cassia, & Umiltà, 1996) and veridical face images (Macchi Cassia, Turati, & Simion, 2004), as compared to equally complex nonfacelike configurations.

There is little dispute that this initial attentional bias toward faces triggers the development of neurocognitive specialization for this class of stimuli, as it results in biased input being provided to the developing cortical tissues within the ventral visual cortex. As a benefit from their frequent exposure to faces, these cortical circuits become progressively structurally and functionally specialized for the processing of stimuli that fall within the face category (Johnson, 2000). Support for this view comes from studies on the effects of early visual deprivation. These studies show that when the face bias cannot exert its influence from birth, as in individuals born with dense bilateral cataracts, face recognition abilities follow an abnormal developmental trajectory (Geldart, Mondloch, Maurer, de Schonen, & Brent, 2002; Le Grand, Mondloch, Maurer, & Brent, 2001).

However, controversy remains about the basis of the newborns' face bias. According to the model of the development of face processing originally proposed by Johnson and Morton (1991) and more recently updated by Johnson (2005), newborns' attention is triggered by faces due to the rapid and automatic activation of a face-configuration detector (i.e., Conspec). This face detector is supported by a fast subcortical processing route operating on low-spatial frequencies, which selectively responds to stimuli consisting of a raised surface with darker areas corresponding to the approximate locations of eyes

---

This research was supported by a grant from the Ministero dell'Università e della Ricerca (No. 2005119101) and a grant from the University of Milano-Bicocca (Fondi di Ateneo per la Ricerca 2004). The authors are deeply indebted to Dr. B. Dalla Barba and the nursing staff at the Pediatric Clinic of the University of Padova and Dr. M.E.. Sotti and the nursing staff at the Casa di Cura of Abano Terme for their collaboration. The authors also thank Dana Kuefner for her thoughtful contribution to the creation of stimuli for Experiment 5, and Stefano Vezzani for his advice concerning the adult visual perception literature.

Correspondence concerning this article should be addressed to Viola Macchi Cassia, Department of Psychology, University of Milano-Bicocca, via dell'Innovazione 10, U9, 20126 Milano, Italy. Electronic mail may be sent to viola.macchicassia@unimib.it.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7904-0001

and mouth. An alternative view, which has recently received much attention within the field, proposes that newborns' face preference arises from a number of domain-general perceptual biases that drive newborns' attention toward certain structural properties of visual stimuli (Simion, Macchi Cassia, Turati, & Valenza, 2001, 2003; Turati, 2004). It has been proposed that these nonspecific biases may likely derive from the functional properties of the newborn's visual system and have a priori nothing to do with faces. Rather, they are domain relevant, allowing newborns to successfully select faces among other nonfacelike stimuli within their species-typical environment.

This claim derives mainly from two lines of reasoning. First, it has been shown that in addition to facedness, newborns manifest spontaneous preferences for other structural properties of visual stimuli (Farroni, Valenza, Simion, & Umiltà, 2000; Simion, Valenza, Macchi Cassia, Turati, & Umiltà, 2002; Slater, Earle, Morison, & Rose, 1985). Therefore, facedness is not the only stimulus dimension that is capable of triggering infants' attention at birth. Second, faces themselves can be described as a collection of perceptual properties. Therefore, the possibility exists that some, if not each of these properties, plays a role in prompting the newborns' attentional response toward facelike stimuli.

Until now, research has focused on one of the structural properties embedded in faces, namely, "top-heaviness." Faces can be described as top-heavy stimuli because they display a greater amount of patterning—the eyes and eyebrows—in the upper portion, and a lesser amount—the mouth—in the lower portion. A series of studies have shown that when embedded in geometric nonfacelike stimuli, the top-heavy property is capable of inducing a preferential visual response in newborns (Simion et al., 2002). This preferential response has been related to a difference in sensitivity between the upper and the lower visual field, similar to the upper visual field advantage observed in adults (Rubin, Nakayama, & Shapley, 1996; Skrandies, 1987), which would render top-heavy patterns more easily detectable than other stimuli. The key finding from these studies, obtained using both schematic (Turati, Simion, Milani, & Umiltà, 2002) and veridical (Macchi Cassia et al., 2004) face images, is that newborns attend equally long to a face as to a perceptually similar nonface stimulus that shares with the face a top-heavy distribution of the features within the outer contour. This demonstration that the facelike arrangement of the inner features displayed by the face did not selectively affect newborns' visual behavior was taken as evidence that what is classically interpreted as a specific inborn preference for the face geometry is likely the result of a more general preference for top-heavy visual objects. Indeed, stimuli within this category include but are not limited to faces.

In the present study, we investigated the contribution to the face preference phenomenon of a second configural visual property included in faces, which we refer to as "congruency." Congruency can be defined, in any given pattern, by the presence of a congruent or corresponding relationship between the shape and the orientation of the bounded area delimiting the pattern and the spatial disposition of the included features. Although the congruency property could be much easily detected in abstract, geometric configurations than in realistic images of real-life visual objects, faces indeed can be described as congruent stimuli. Despite within-category variability in perceptual appearance, faces on average are delimited by a mono-oriented frame, which appears wider in the upper part where the hair and ears are located and narrower in the lower part where the chin and the neck are located. Therefore, given the typical spatial arrangement of the eyes and mouth, faces display a greater number of features in the widest portion and only one feature in the narrowest part. Following the same logic used in earlier studies investigating the contribution of the top-heavy bias to newborns' face preference, we reasoned that if the congruency property is capable of inducing a preferential response when embedded in nonfacelike configurations, then it could be argued that this property also plays a role in inducing newborns' preference for faces.

The hypothesis that congruent visual configurations may be preferred at birth over noncongruent patterns appears reasonable in light of three lines of research. First, there is considerable evidence that newborns are highly sensitive to configural–holistic properties emerging from the interrelations between the component parts of the stimuli. For instance, they can perceive the invariance of the spatial relationship between single features, which vary in their absolute position within an array (Antell, Caron, & Myers, 1985). Also, newborns are able to group separate sets of elements according to Gestalt principles (Farroni et al., 2000), and they find configural and global cues in hierarchical patterns more easily detectable than featural and local information (Macchi Cassia, Simion, Milani, & Umiltà, 2002). Second, based on Gestalt theories of visual perception (Garner, 1974; Palmer, 1991), in comparison to noncongruent configurations, congruent configurations provide a better fit for the criteria of figural simplicity and regularity that render visual patterns more easily and economically processed and represented by the human perceptual system. Because newborns have been shown to perceive and organize visual arrays according to Gestalt principles, such as lightness similarity (Farroni et al., 2000) and common motion (Valenza & Bulf, 2007), it seems reasonable to hypothesize that newborns may also be sensitive to other stimulus dimensions that contribute to figural goodness, such as symmetry, repetition, and regularity. Each of these dimensions is maximally present in congruent visual configurations. Third, preliminary findings from an exploratory visual preference study showed that when newborns were presented with a 180° rotated version of the schematic facelike and nonfacelike configurations used in earlier face preference studies, they did not show a preferential response for either of the two stimuli ($p = .2$; Macchi Cassia, Valenza, and Simion, unpublished data; Figure 1). A possible interpretation of this finding is that each of the stimuli contained a relevant property, that is, top-heaviness or congruency, and that these two properties competed in triggering newborns' attention. Indeed, being upside down, the schematic face displayed a congruent relationship between the orientation of the outer border and the spatial locations of the included elements (i.e., two elements were located in the widest portion of the border) in the absence of a top-heavy arrangement of those elements (i.e., two elements were located in the lower part of the triplet and one in the upper part). In contrast, the rotated nonfacelike stimulus displayed a noncongruent relation between the border and the inner elements in the presence of a top-heavy arrangement of those elements. Therefore, newborns may have failed to show a preference for either of the two stimuli as a consequence of the competing properties embedded in the stimuli.

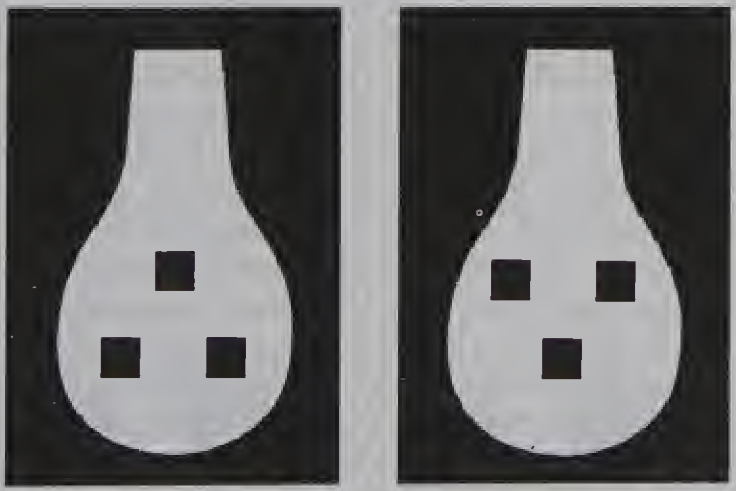


Figure 1. The 180° rotated version of the schematic facelike and nonfacelike configurations used in the exploratory visual preference study.

Note. Newborns did not show a preferential response for either of the two stimuli (Macchi Cassia, Valenza, and Simion, unpublished data).

In the present study, we aimed to test the role played by the perceptual property of congruency in newborns' face preference. To do so, we used the same approach that was used in previous studies investigating the role of the top-heavy property in newborns' face preference. The rationale of those studies was that of altering the spatial integrity of the facelike configuration so as to selectively break one of the structural rules embedded in faces and testing the effect of each manipulation on newborns' visual preferences. The same systematic approach of decomposing the structural properties defining facedness was applied in the present study, where we measured newborns' attentional responses to compound geometric stimuli in which we independently manipulated the facelike versus nonfacelike arrangement of the inner elements and the congruent versus noncongruent relation of the elements with the shape of the external contour. It is important to note that in the facelike stimuli used in Experiments 1 through 4, the facedness and the top-heavy property were inextricably mixed, and therefore, we refer to those stimuli as *facelike* and make reference to *facedness* as the crucial property defining those stimuli. In the stimuli used in Experiment 5, the number and shape of the inner elements and the spatial integrity of the facelike configuration were altered. Therefore, we refer to those stimuli as nonfacelike *top-heavy* and make reference to top-heaviness as the crucial property defining them.

In the five reported experiments, newborns were tested using a preferential looking paradigm. The specific aims of the experiments were (a) to determine whether the congruency property elicits a preference when it is embedded in nonfacelike stimuli (Experiment 1); (b) to test the relative strength of congruency and facedness in eliciting newborns' preference (Experiment 2); (c) to test the hypothesis of an additive model, where facedness and congruency contribute to the same extent to trigger newborns' attention toward faces (Experiments 3 and 4); and (d) to extend the validity of the additive model when congruency is added to top-heaviness in nonfacelike stimuli displaying more elements in the upper portion (Experiment 5).

Experiment 1

The aim of the first experiment was to provide evidence as to whether newborns show a spontaneous visual preference for congruent over noncongruent nonfacelike visual configurations. To test this hypothesis, newborns were presented with two compound stimuli comprising three black squares embedded in a triangular-shaped area. The orientation of the triangular-shaped areas was either congruent or noncongruent with the spatial arrangement of the inner elements (Figure 2). The spatial arrangement of the three squares was analogous to that used in the inverted nonfacelike stimuli employed in the studies of Johnson and Morton (1991) and Valenza et al. (1996), showing newborns' preference for schematic faces.

Method

Participants. The participants were 12 (5 males) healthy, full-term infants, 25–73 hr old, recruited from the maternity wards of the Pediatric Clinic at the University of Padova and the Gynecology Unit at the Hospital of Abano Terme. They were middle-class infants, and 99% of them were Caucasian and 1% was Asian. Two additional infants were tested, but one did not complete testing due to fussiness and one was excluded from the data analysis due to position bias (i.e., looking more than 80% of the time in one direction). Infants were tested only in an awake, alert state, after the parents had given their informed consent.

Stimuli. Each stimulus comprised three black square elements that were 2.2 cm (4.2°) on each side and located into a white triangular-shaped area that was 13.5 cm (24.2°) in width and 15 cm (26.6°) in height. The squared elements were arranged in a nonfacelike disposition, with two elements in the lower part and one in the upper part, so as to create a triangle with one vertex pointing up. The distance between the three elements was 1 cm (1.9°). The two stimuli were identical except for the orientation of the white area, which paralleled the orientation of the triangle created by the three inner elements with one vertex pointing up in the congruent stimulus, and rotated 180° so that the vertex was pointing down in the noncongruent stimulus (Figure 2).

Apparatus. The infant sat on a student's lap, in front of a black panel, at a distance of about 30 cm. The panel had two square holes where the black screens of two computer monitors appeared. The infant's eyes were aligned with a red flickering light emitting diode (LED), located in the center of the screen. The LED was used to attract the infant's gaze at the start of both the habituation and the preference test phases, subtended about 2° of a visual angle and, when turned on, blinked at a rate of 300 ms. Stimuli were projected bilaterally at a distance of approximately 7 cm from the central LED. To prevent interference from irrelevant stimuli, two black panels were placed on both sides of the infant to limit peripheral vision.

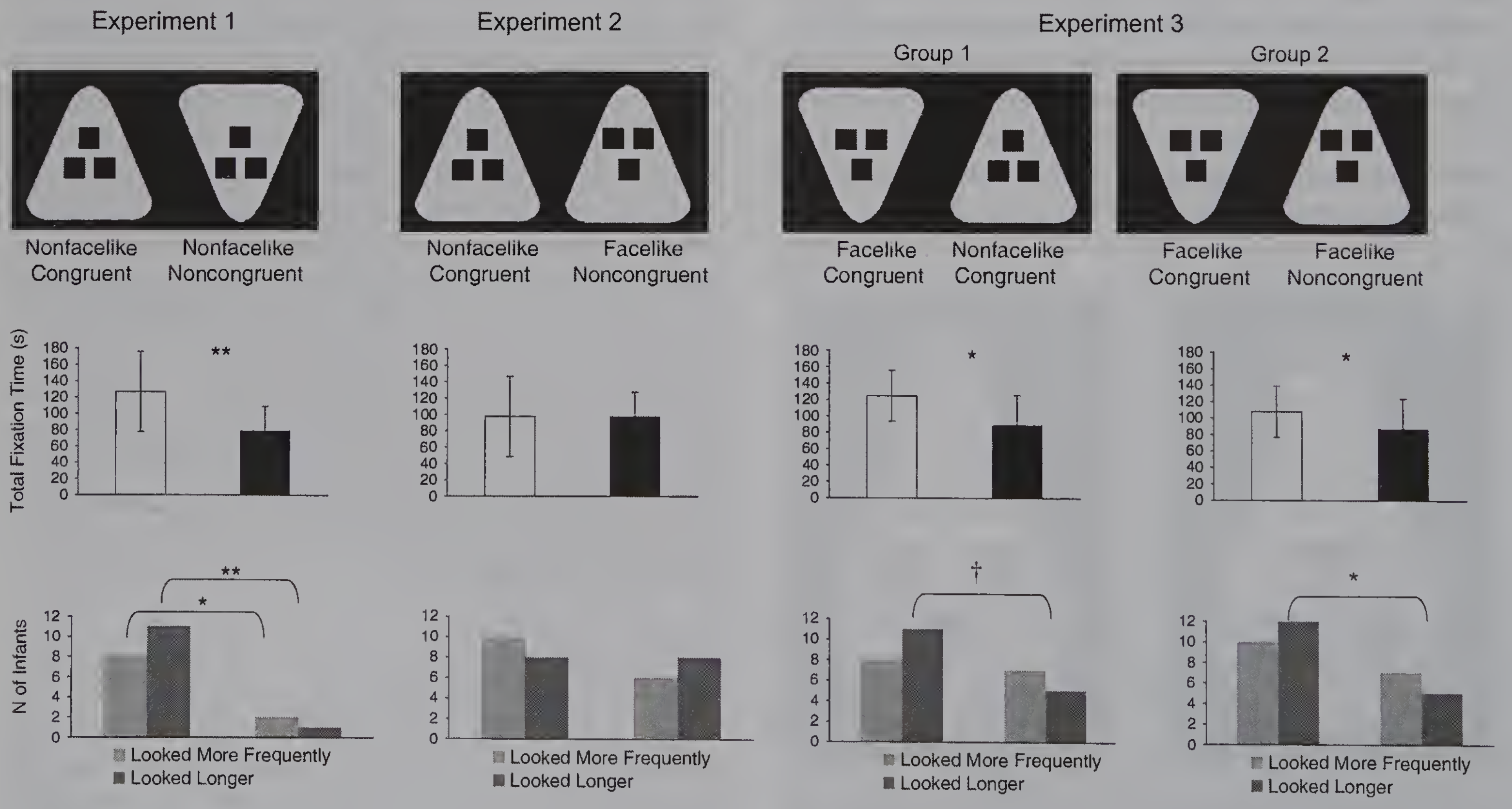


Figure 2. Stimulus pairs and results from Experiments 1, 2, and 3.

Note. For each experiment, two graphs are reported. The first graph shows the total fixation time toward each stimulus, with error bars indicating standard deviations. The second graph shows the number of newborns who oriented more frequently and/or looked longer toward each stimulus. For each graph, the *p* value is reported for the comparison between the stimuli.

$^*p < .05$. $^{**}p < .01$. $^\dagger p = .07$.

Procedure. Each trial began with the central flickering LED. As soon as the infant fixated the LED, an experimenter started the sequence of trials by pressing a key on the computer keyboard. This automatically turned off the central LED and activated the stimuli on the screens. Stimuli remained on as long as the infant fixated one of them (i.e., infant control procedure). When the infant shifted his gaze from the display for more than 10 s, the experimenter turned off the stimuli and the central LED automatically turned on. All infants were presented with two trials, in which the position of the stimuli was counterbalanced. All testing sessions were video recorded. Videotapes of eye movements throughout the trial were subsequently analyzed frame by frame to the nearest 40 ms by two coders, both unaware of the stimuli presented. The coders recorded number of orienting responses toward the stimuli and total fixation times, that is, the sum of all fixations (Cohen, 1972, 1973). Intercoder agreement (Pearson correlation) was .99 for total fixation time and .91 for discrete number of looks.

Results

Preliminary analyses of variance (ANOVAs) showed no significant main effects or interactions involving the within-subjects factor of trial (first or second) for either dependent variable. Therefore, two separated *t* tests for dependent samples were performed to compare number of orienting responses and total fixation time to the congruent and the noncongruent stimulus. Newborns showed a nonsignificant trend to orient more frequently to the congruent pattern ($M = 15.7$, $SD = 5.4$) rather than to the noncongruent pattern ($M = 12.6$, $SD = 6.3$), $t(11) = 1.76$, $p = .106$. This trend reached statistical significance for total fixation times, which were 22% longer to the congruent stimulus ($M = 126.8$ s, $SD = 49.2$) than to the noncongruent stimulus ($M = 78.9$ s, $SD = 30.0$), $t(11) = 3.42$, $p = .006$ (both tests were two-tailed; Figure 2).

Finally, we examined the data for individual infants and compared the number of infants who oriented more frequently and/or looked longer at the congruent stimulus than at the noncongruent stimulus. The comparisons were performed by means of two different statistical tests: a binomial test and a nonparametric Wilcoxon signed-rank test, which was used in consideration of the small sample numerosity. Eight of 10 infants (2 infants were not included in the count because they produced the same number of looks to the two stimuli) made a greater number of looks to the congruent stimulus (8 vs. 2, binomial test $p = .044$ and Wilcoxon test $p = .092$), and 11 of 12 infants looked longer at the congruent stimulus (11 vs. 1, binomial test $p = .003$ and Wilcoxon test $p = .012$; Figure 2). The difference in looking times for the two stimuli ranged from 11.9 to 147.7 s.

Discussion

The results showed that the presence of a congruent relation between the shape and the orientation of the bounded area delimiting the pattern and the spatial disposition of the included features is one of the stimulus properties that control newborns' visual preferences. The two stimuli used in the present experiment contained the same amount of visual information, displayed the same symmetrical configuration, and neither resembled a face in that the spatial disposition of the inner elements did not simulate the face geometry. Rather, the stimuli differed exclusively in the orientation of the triangular-shaped contour and, therefore, in the congruent versus noncongruent relation between the contour and the triplet of inner elements. Therefore, congruency can be regarded as the crucial factor that produced the observed preference.

This finding proves particularly relevant because it provides the first demonstration of the presence in newborns of a spontaneous attentional bias toward a configural/relational perceptual property, that is, a property emerging from the spatial interrelations among the stimulus components (Kimchi, 1992). Most crucially, the demonstration that the congruent relation between the disposition of the inner elements and the shape of the external contour is capable of triggering newborns' attention when embedded in nonfacelike configurations raises the possibility that this same property may also play a role in triggering newborns' attention toward facelike stimuli. The spontaneous preference for the congruency property may be a contributing factor that, interacting with other factors, gives rise to the newborns' face preference phenomenon. This possibility was explored in the following experiments.

Experiment 2

To explore the strength of the congruency property in prompting newborns' attentional responses to facedness, in Experiment 2, the nonfacelike congruent stimulus preferred in Experiment 1 was contrasted to a noncongruent stimulus displaying a facelike arrangement of the inner elements (Figure 2). The facelike spatial arrangement of the triplet of elements in the noncongruent stimulus was analogous to that used in the upright facelike stimuli employed in the studies of Johnson and Morton (1991) and Valenza et al. (1996), showing newborns' preference for schematic faces. A preference for either of the two stimuli would indicate which of the two stimulus properties (i.e., facedness vs. congruency) is more influential in driving newborns' attention. Crucially, if newborns would not manifest a preference, this would imply that none of the two properties is more powerful than the other in triggering attentional responses at birth.

Method

Participants. The final sample consisted of 16 (10 males) healthy, full-term, middle-class, Caucasian infants, 24–79 hr old. Four additional infants were tested but not included in the sample because of fussiness ($n = 2$) or position bias ($n = 2$).

Stimuli. The stimuli in this experiment were the congruent nonfacelike pattern presented in Experiment 1 and a noncongruent facelike pattern, which was matched to the congruent pattern for the orientation of the white area and differed exclusively for the spatial disposition of the three inner elements (Figure 2). In both stimuli, the triangular-shaped white area had one vertex pointing up. In the facelike stimulus, two elements of the triplet were located in the upper part of the configuration and one in the lower part, whereas in the nonfacelike stimulus, the triplet of elements was rotated 180° so that two elements appeared in the lower part and one in the upper part.

Apparatus and procedure. The same apparatus and procedure from Experiment 1 were used.

Results

Preliminary ANOVAs showed no significant main effects or interactions involving the within-subjects factor of trial (first or second) for either dependent variable. Two paired-sample *t* tests were performed to determine whether newborns' number of orienting responses and total fixation time differed for the two stimuli presented. The difference was not significant for either dependent variable (Figure 2). Newborns looked 14.9 ($SD = 6.7$) times at the noncongruent facelike stimulus and 16.2 ($SD = 9.3$) times at the congruent nonfacelike stimulus, $t(15) = 1.26, p = .226$. Similarly, they spent 96.2 s ($SD = 42.0$) and 97.5 s ($SD = 60.9$) looking at the noncongruent and congruent stimuli, respectively, $t(15) = 0.07, p = .946$ (both tests were two-tailed). Finally, examination of the data for individual infants revealed that the number of infants who produced a greater number of looks and/or looked longer at the noncongruent facelike stimulus did not differ from the number of infants who oriented more frequently toward (6 vs. 10, binomial test $p = .122$ and Wilcoxon test $p = .242$) and/or fixated longer (8 vs. 8, binomial test $p = .196$ and Wilcoxon test $p = .717$) the congruent nonfacelike stimulus (Figure 2). The difference in looking times for the two stimuli ranged from 14.8 to 171.7 s.

Discussion

Although nonsignificant results should be interpreted with caution, these findings seem to suggest that facedness per se was not the only factor affecting newborns' preferences in the present experiment. The absence of a preference for either stimulus likely indicates that facedness and the configural property of congruency contributed to the same extent to drive newborns' preferences. These findings confirm and extend the null results obtained in the exploratory study with 180° rotated schematic facelike and nonfacelike stimuli (Macchi Cassia, Valenza, and Simion, unpublished data; Figure 1), providing further evidence that when facedness and congruency are pitted against each other, these two stimulus properties evoke a competition for newborns' attentional resources. The result is a failure for newborns to show a preference for either stimulus.

The demonstration that the preference for the congruent stimulus was not overcome by the preference for the facelike stimulus lands support to the hypothesis that both stimulus properties—congruency and facedness—may play a role in inducing newborns' preference for faces, given that both are properties defining the face stimulus category. More specifically, the lack of any significant result in Experiment 2 might support the hypothesis that the attentional biases toward facedness and congruency produce a cumulative effect on newborns' visual preferences according to an additive model, in which both biases would exert an equally critical influence on newborns' preferential responses.

Experiment 3

The aim of Experiment 3 was to provide a direct test of the additive role of facedness and congruency in inducing newborns' preferences. Two different groups of infants were tested, each presented with a pair of stimuli that were equated for one of the two properties and differed for the presence versus absence of the other property. For Group 1, both stimuli within the pair preserved a congruent relation between the inner elements and the outer frame, and only one of the stimuli displayed the facelike arrangement of the inner elements that defines facedness. For Group 2, both stimuli displayed the facedness property, and only one of the two displayed a congruent relation between the inner elements and the outer frame (Figure 2). Based on the evidence provided by Experiment 2, we hypothesized that the attentional biases toward the two stimulus properties of facedness and congruency may exert a cumulative influence on newborns' visual preferences, according to an additive model. Based on this hypothesis, we predicted that both groups of infants would prefer the stimulus in which both properties were present over the stimulus in which only one of the two properties was present. The alternative hypothesis, according to which the facelike disposition of the three inner elements that produces facedness, per se, is sufficient to trigger newborns' attention toward faces (Johnson & Morton, 1991), would predict that no preference should be found for Group 2, whereas a preference for the facelike congruent stimulus should be found for Group 1.

Method

Participants. The final sample consisted of 33 (14 males) healthy, full-term infants, 30–65 hr old, randomly assigned to two groups (16 in Group 1 and 17 in Group 2). Participants were middle-class infants, and 92% of them were Caucasian, 1% African, and 7% Asian. Ten additional infants were tested but not included in the sample due to position bias ($n = 1$) or experimenter error ($n = 1$), or because they did not complete testing due to fussiness ($n = 8$).

Stimuli. Two different pairs of stimuli of the same type as those used in Experiments 1 and 2 were employed, one for each group of participants. Each pair comprised a congruent facelike stimulus in which both the stimulus properties of facedness and congruency were present, and a stimulus similar to the previous one in which the facedness property (Group 1) or the congruency property (Group 2) was selectively removed. Therefore, in the first pair of stimuli (Group 1), the congruent facelike stimulus was contrasted with a congruent nonfacelike stimulus, whereas in the second pair (Group 2), the congruent facelike stimulus was contrasted with a noncongruent facelike stimulus (Figure 2). As in Experiments 1 and 2, the facelike versus nonfacelike appearance of the stimuli was produced by a 0° versus 180° rotation of the triplet of inner elements, and the congruent versus noncongruent nature of the stimuli was a product of the rotation of the triangular-shaped white area in which the elements were comprised.

Apparatus and procedure. The apparatus and procedure were the same as those used in Experiment 1.

Results

Preliminary ANOVAs showed a significant main effect of the within-subjects factor of trial (first or second) for total fixation times, $F(1, 31) = 10.27$, $p = .003$, due to newborns producing longer fixation times in the first trial compared to the second trial. However, no significant interactions involving the factor of trial were found for either dependent variable. Therefore, two ANOVAs with group (Groups 1 and 2) as a between-subjects factor and stimulus properties (two properties and one property) as a within-subjects factor were performed, one for number of orienting responses and the other for total fixation time. No significant effects emerged for number of orienting responses. Newborns in Group 1 looked 13.1 ($SD = 4.4$) times at the congruent facelike stimulus and 11.8 ($SD = 5.0$) times at the congruent nonfacelike stimulus, $t(15) = 0.86$, $p = .401$, and newborns in Group 2 looked 11.7 ($SD = 6.2$) times at the congruent facelike stimulus and 10.4 ($SD = 4.4$) times at the noncongruent facelike stimulus, $t(16) = 0.91$, $p = .379$. For total fixation time, there was a main effect of stimulus properties, $F(1, 31) = 10.27$, $p = .003$, due to fixation times to the congruent facelike stimulus being longer than fixation times to the second stimulus of the pair for each group. For Group 1, fixation times were 18% longer for the congruent facelike stimulus ($M = 124.7$ s, $SD = 40.5$) than for the congruent nonfacelike stimulus ($M = 89.0$ s, $SD = 39.7$), $t(15) = 2.29$, $p = .037$. For Group 2, the facelike congruent stimulus was fixated 11% longer ($M = 108.0$ s, $SD = 31.1$) than the facelike noncongruent stimulus ($M = 87.3$ s, $SD = 36.6$), $t(16) = 2.36$, $p = .031$ (all tests were two-tailed; Figure 2).

Examination of the data for individual infants revealed that for Group 1, the number of infants who produced a greater number of looks at the congruent facelike stimulus did not differ from the number of infants who oriented more frequently toward the congruent nonfacelike stimulus (8 vs. 7, binomial test $p = .196$ and Wilcoxon test $p = .348$), with one infant not included in the count because he or she made the same number of looks to the two stimuli. The number of infants who looked longer at one or the other stimulus was only marginally significant (11 vs. 5, binomial test $p = .067$ and Wilcoxon test $p = .063$). For Group 2, 10 of 17 infants oriented more frequently toward the congruent facelike stimulus compared to the noncongruent facelike stimulus (10 vs. 7, binomial test $p = .148$ and Wilcoxon test $p = .651$), and 12 of 17 looked longer at this same stimulus (12 vs. 5, binomial test $p = .047$ and Wilcoxon test $p = .035$; Figure 2). For both groups, the differences in looking times for the two stimuli ranged from 3.8 to 111.3 s.

Discussion

Results from Experiment 3 showed that newborns manifested a spontaneous visual preference for stimuli in which both the facedness and the congruency properties were present over stimuli that comprised only one of these two properties. Specifically, evidence from Group 1 showed that the strength of congruency in triggering newborns' attention increased when this property was accompanied by facedness. Evidence from Group 2 showed that facedness per se was not sufficient to preferentially attract newborns' attention and that the attention-triggering value of the facelike arrangement of the inner elements was enhanced when the elements maintained a congruent relation with the external frame. This finding does not support the hypothesis of the existence at birth of a Conspec-like mechanism, which would be automatically activated by stimuli displaying the unique geometry of the face, defined by the correct relative location of the internal features for the eyes and mouth (i.e., facedness; Johnson & Morton, 1991; Johnson, 2005). Rather, the present findings are in line with recent claims that a number of attentional biases toward nonspecific stimulus properties combine to produce the face preference and demonstrate that the congruency bias is one of those biases (Simion et al., 2001, 2003).

Results from Experiment 3 demonstrate that the attentional biases toward facedness and congruency have an additive effect in driving newborns' preferences. The finding that the congruent facelike stimulus was preferred over both the congruent nonfacelike (Group 1) and the noncongruent facelike (Group 2) stimuli provide further support to the interpretation of the null results of Experiment 2. In fact, the two stimulus properties of facedness and congruency were capable of triggering newborns' attention to the same extent in the two groups, and their cumulative influence on newborns' visual preferences followed an additive rule. This finding provides support for the argument that both attentional biases toward facedness and congruency are involved in the preferential responses that newborns manifest for schematic faces.

One final note of caution concerns the evidence emerging from Group 2, showing that facedness did not exert an all-or-nothing influence on newborns' attention. This evidence derives from the comparison between two facelike stimuli, one displaying a congruent relation between the inner elements and the outer frame and the other in which the congruency property was not only missing but also intentionally violated by the misorientation of the outer frame. It is possible that the reduction of the figural goodness of the stimulus induced by this violation, rather than the cumulative effect produced by both the facedness and the congruency properties of the comparison stimulus, may have driven the preference for the congruent facelike stimulus observed in Group 2. If true, it could be hypothesized that if the congruent facelike stimulus was compared to a facelike stimulus in which the congruency property was neutralized, rather than violated, the facedness property would trigger newborns' attention to the same extent in both stimuli. This possibility was tested in Experiment 4.

Experiment 4

Experiment 4 was aimed at testing whether the preference observed in Group 2 of Experiment 3 would still be obtained when the congruent facelike stimulus is compared to a facelike configuration in which the triplet of inner elements is surrounded by a nonoriented, squared-shaped contour (Figure 3). If the congruent facelike stimulus was preferred in Experiment 3 because it included the two preferred properties of facedness and congruency, this preference should be maintained in Experiment 4. If, in contrast, the preference for the facelike congruent stimulus in Experiment 3 was a by-product of the reduced perceptual appeal produced by the violation of the congruency property in the stimulus to which it was compared, no preference should be expected in the present experiment.

Method

Participants. The final sample consisted of 15 (9 males) healthy, full-term, Caucasian infants, 25–80 hr old. Two additional infants were tested but not included in the sample due to position bias ($n = 1$) or fussiness ($n = 1$).

Stimuli. The stimuli in this experiment were the facelike congruent configuration presented in Experiment 3 and a facelike neutral stimulus in which the three squared elements were comprised within a white rectangular-shaped area whose width (24.2°) and height (26.6°) were the same as those of the triangular-shaped area surrounding

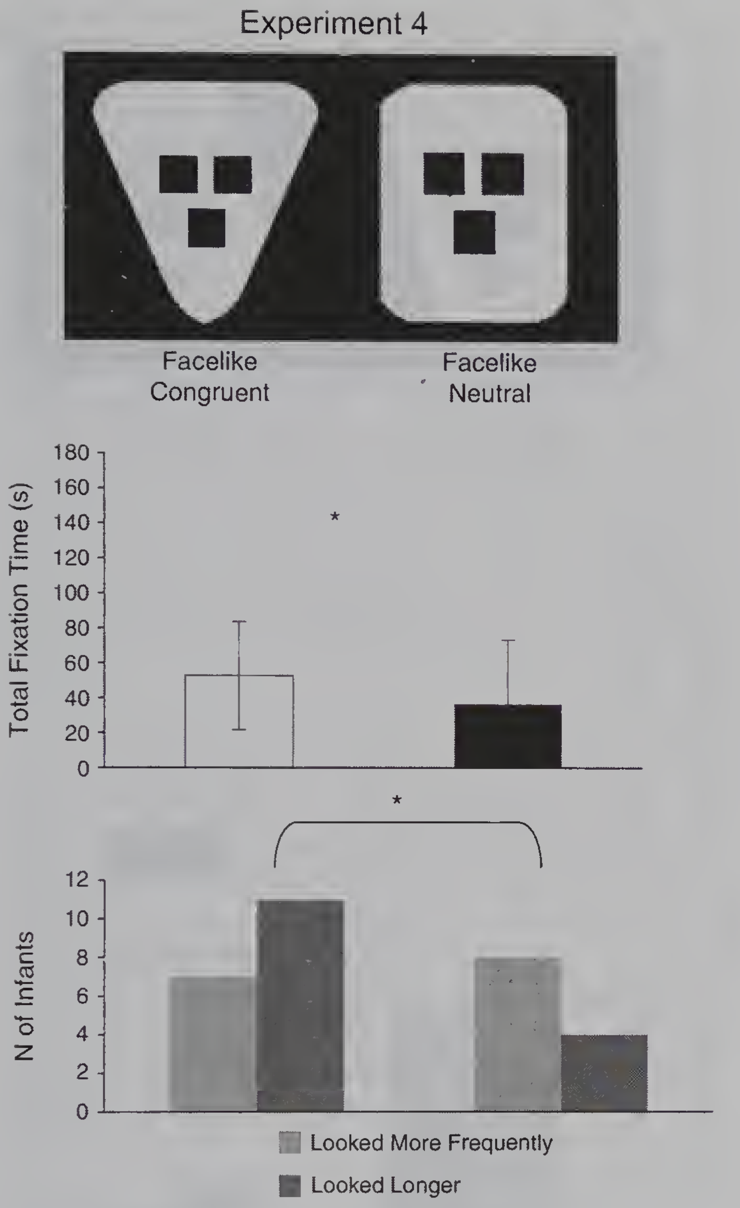


Figure 3. Stimulus pairs and results from Experiment 4.
Note. The first graph shows the total fixation time toward each stimulus and the second shows the number of newborns who oriented more frequently and/or looked longer toward each stimulus. The p values refer to the comparison between the stimuli, and the error bars indicate standard deviations.
$^*p < .05$.

the inner elements in the facelike congruent stimulus (Figure 3).

Apparatus and procedure. The apparatus and procedure were the same as those used in Experiment 1.

Results

Preliminary ANOVAs showed a significant main effect of the within-subjects factor of trial (first or second) for total fixation times, $F(1, 14) = 10.08$, $p = .007$, due to fixation times being longer during the first trial compared to the second trial, but no significant interactions involving this factor were found for either dependent variable. Therefore, two separated t tests for dependent samples were performed to compare number of orienting responses and total fixation time to the two stimuli presented. No difference was found for number of orienting responses. Newborns looked 15.5 ($SD = 5.8$) times at the facelike congruent stimulus and 14.1 ($SD = 6.6$) times at the facelike neutral stimulus, $t(14) = 0.70$, $p = .495$. In contrast, fixation times were 20% longer to the facelike congruent stimulus ($M = 52.7$ s, $SD = 24.9$) than to the facelike neutral stimulus ($M = 36.6$ s, $SD = 18.2$), $t(14) = 2.31$, $p = .037$ (Figure 3).

Examination of the data for individual infants showed that almost half of the infants in the sample oriented more frequently toward the facelike congruent stimulus (7 vs. 8, binomial test $p = .196$ and Wilcoxon test $p = .733$), whereas two thirds of the infants looked longer at this same stimulus (11 vs. 4, binomial test $p = .042$ and Wilcoxon test $p = .061$; Figure 3). The differences in looking times for the two stimuli ranged from 1.6 to 68.0 s.

Discussion

The results showed that the preference for the congruent facelike stimulus persevered when this stimulus was contrasted to a facelike stimulus in which the congruency property was neutralized by the presence of a nonoriented external contour. This finding demonstrates that the preference for the congruent facelike stimulus over the noncongruent facelike stimulus observed in Experiment 3 was truly due to the cumulative effect of the facedness and congruency properties. Moreover, this finding replicates and extends the results of Experiment 3, showing that the facedness property is capable of eliciting a preferential response only when the congruency property is concurrently present within the stimulus.

By replicating the findings of Experiment 3, the results of Experiment 4 provided further evidence for the additive effect exerts by congruency and facedness on newborns' visual preferences. The demonstration that the attentional biases toward these two stimulus properties produce a cumulative effect on newborns' visual preferences according to an additive model is in agreement with the claim that newborns' face preference derives from the cumulative action of a set of attentional biases toward different nonspecific stimulus properties that faces share with other nonface stimuli, rather than from one single specific bias toward facedness (Simion et al., 2001, 2003). However, this claim would receive much stronger support by the demonstration that the same cumulative effect on newborns' preferences could be found with congruent stimuli in which facedness is not confounded with top-heaviness. In fact, it is important to highlight that the nonspecific structural property of top-heaviness was inherently embedded in the facelike stimuli used in Experiments 3 and 4. Therefore, top-heaviness may have played a critical role in triggering newborns' preferences toward these stimuli. Using top-heavy stimuli displaying a nonfacelike disposition of the inner elements, in Experiment 5, we tested the hypothesis that the two perceptual properties of congruency and top-heaviness produce a cumulative effect on newborns' preferences.

Experiment 5

Experiment 5 was aimed at extending the results obtained for Group 2 of Experiment 3 to top-heavy stimuli that do not display the facelike arrangement of the inner elements, which defines facedness. Specifically, the goal was to provide direct evidence concerning the additive effect of the top-heavy bias and the congruency bias in driving newborns' preferences. Newborns were presented with a pair of compound stimuli comprising five elements, which were arranged so as to create a top-heavy nonfacelike configuration. Specifically, the elements were more numerous in the upper part compared to the lower part but not disposed in correct location for the eyes and mouth, as it was in the schematic facelike stimuli used in Experiments 2 through 4. Although being equated for the top-heavy configuration of the elements, the two stimuli differed for the orientation of the white area surrounding them, which was congruent with the spatial arrangement of the elements in one stimulus and incongruent with those elements in the second stimulus (Figure 4).

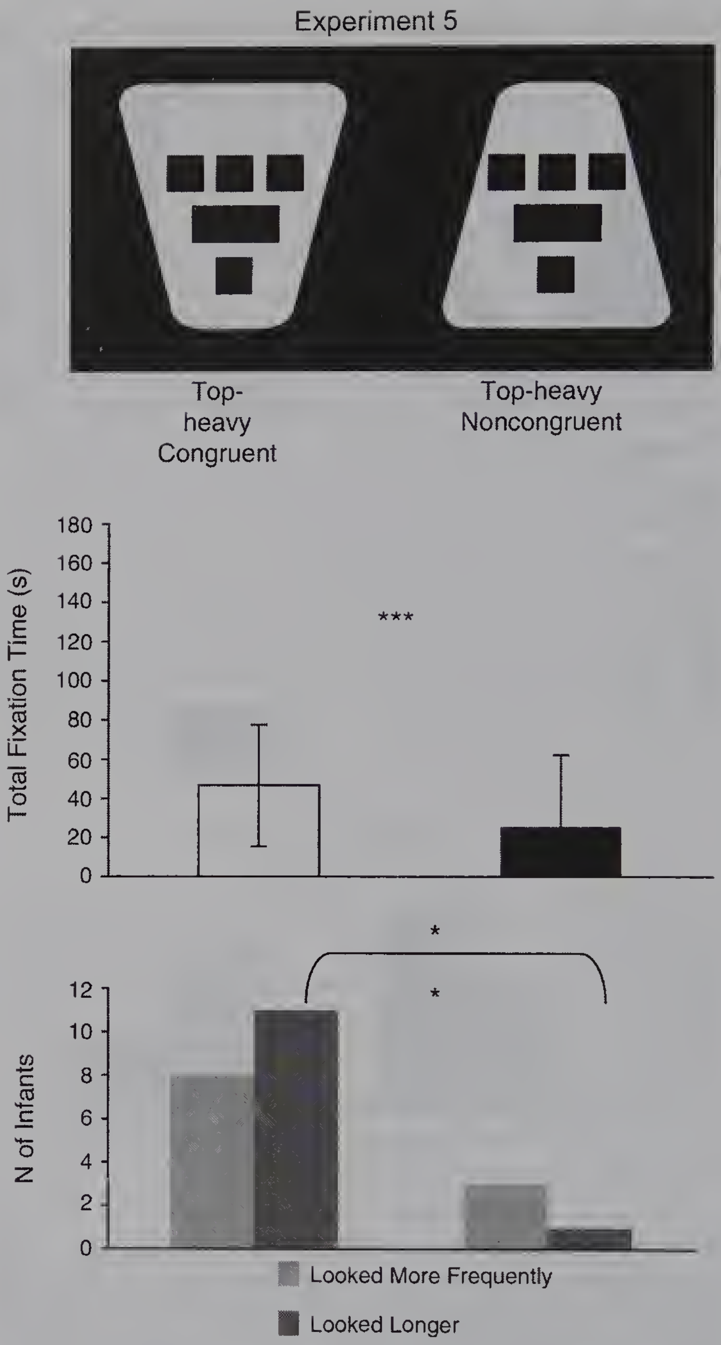


Figure 4. Stimulus pairs and results from Experiment 5.
Note. The first graph shows the total fixation time toward each stimulus and the second shows the number of newborns who oriented more frequently and/or looked longer toward each stimulus. The p values refer to the comparison between the stimuli, and the error bars indicate standard deviations.
$^{*}p < .05$. $^{***}p < .001$.

Method

Participants. The final sample consisted of 12 (7 males) healthy, full-term newborn infants, 27–82 hr old. Participants were middle-class infants, and 98% of them were Caucasian and 2% Asian. Two additional infants were tested but not included in the sample due to fussiness.

Stimuli. Each stimulus comprised four black square elements of the same size as those composing the stimuli used in Experiments 1 through 4 and one rectangular-shaped element that was 2.2 cm (4.2°) high and 5.9 (11.2°) wide. The distance between the elements was 1 cm (1.9°). Three of the four squared elements were located in the upper part of the stimulus, and only one was located in the lower part, with the rectangular-shaped element located in the middle, so as to create a top-heavy configuration. The elements were located into a white area, which differed in shape but not in size from that used in Experiments 1 through 4. The orientation of the white area paralleled the top-heavy distribution of the elements in the congruent stimulus and rotated 180° on the noncongruent stimulus (Figure 4).

Apparatus and procedure. The apparatus and procedure were the same as those used in Experiment 1.

Results

Preliminary ANOVAs showed no significant main effects or interactions involving the within-subjects factor of trial (first or second) for either dependent variable. Therefore, two paired-sample t tests were performed to determine whether newborns' number of orienting responses and total fixation time differed for the two stimuli presented. Newborns showed a marginally significant trend to orient more frequently to the top-heavy congruent stimulus ($M = 10.1$, $SD = 3.1$) rather than the top-heavy noncongruent stimulus ($M = 8.2$, $SD = 3.5$), $t(11) = 1.88$, $p = .089$. This trend reached statistical significance for total fixation times, which were 29% longer to the top-heavy congruent stimulus ($M = 46.6$ s, $SD = 12.7$) than to the noncongruent stimulus ($M = 25.6$ s, $SD = 9.8$), $t(11) = 5.21$, $p < .001$ (both tests were two-tailed; Figure 4).

Examination of the data for individual infants showed that only two thirds of the infants in the sample oriented more frequently toward the top-heavy congruent stimulus (8 vs. 3, binomial test $p = .081$ and Wilcoxon test $p = .107$), with one infant not included in the count because he or she made the same number of looks to the two stimuli, and almost all the infants looked longer at the same stimulus (11 vs. 1, binomial test $p = .003$ and Wilcoxon test $p = .004$; Figure 4). The differences in looking times for the two stimuli ranged from 4.9 to 40.5 s.

Discussion

The results showed that the top-heavy congruent stimulus was preferred over the top-heavy noncongruent stimulus, thus indicating that the congruency and top-heavy properties have an additive effect on newborns' preferences. When compared to the results obtained for Group 2 of Experiment 3, the findings from the present experiment suggest that the presence of the congruency property is crucial in determining the preference for top-heavy configurations in the same way as it is crucial for the emergence of the preference for schematic facelike configurations.

General Discussion

The results of the present study extend our knowledge concerning the stimulus properties capable of eliciting spontaneous visual preferences at birth (see Slater, 1995, for a review). Evidence from Experiment 1 shows that when two stimuli for which no preference has been shown to exist at birth are equated for the quantity of energy they contain (i.e., number of elements, contrast information), a configural property emerging from the spatial relations among the stimulus components (Kimchi, 1992) can induce a visual preference in newborns. This finding adds to other evidence showing that newborns manifest a preference for stimuli displaying specific distributions of energy within the pattern (i.e., specific structural properties), such as that for horizontal over vertical configurations (Farroni et al., 2000; Slater & Sykes, 1977; Slater et al., 1985) and for top-heavy patterns over bottom-heavy patterns (Simion et al., 2002). Together, these findings demonstrate that newborns are sensitive to stimulus dimensions other than those specified by physical, low-level properties (i.e., contrast information, spatial frequencies; Banks & Ginsburg, 1985; Banks & Salapatek, 1981).

Crucially, these findings are the first to demonstrate that newborns are sensitive to a configural, relational property generated by the interdependency among different features of the stimulus—the shape/orientation of the outer frame and the spatial relation among the inner features. In fact, to prefer the congruent pattern over noncongruent pattern in Experiment 1, and in the other experiments, newborns must not only organize the three squared blobs in a triangular configuration but also perceive the relationship between the triangular configuration of the internal elements and the orientation of the external frame. Therefore, overall the present findings demonstrate that newborns are sensitive to the relation between inner and outer features of the stimulus and appreciate the redundancy and regularity of this relation, independently of the facelike or nonfacelike appearance of the stimulus. Notably, with regard to faces, the presence since birth of the basic ability to extract the perceptual attributes of the relation between the inner and the outer features of the stimulus does not prevent infants from treating these features as independent cues for individual face recognition until around 7 months of age, when holistic processing of internal and external facial features has been found to first emerge (Cashon & Cohen, 2004; Schwarzer, Zauner, & Jovanovic, 2007).

Just as for preferences based on the physical properties of stimuli (Banks & Ginsburg, 1985; Banks & Salapatek, 1981), stimulus preferences based on the energy distribution within the pattern have been interpreted as resulting from constraints on the newborns' visual system, which would render some stimuli more easily detectable and explored than others. For example, newborns' preference for horizontal configurations over vertical configurations has been attributed to eye movements, which at birth tend to be greater in number and size in the horizontal, rather than the vertical direction (Slater & Sykes, 1977). Similarly, as already mentioned, newborns' preference for top-heavy configurations over bottom-heavy configurations has been related to a possible difference in sensitivity between the two vertical hemifields, which would parallel that observed in adults (Rubin, Nakayama, & Shapley, 1996; Skrandies, 1987). Evidence deriving from a neural network simulation study conducted by Acerra, Burnod, and de Schonen (2002) suggests that a similar interpretation based on constraints on the newborns' visual system may also apply to the preference for congruent stimuli over noncongruent stimuli. In comparison to noncongruent patterns, the regular disposition of the internal elements with respect to the shape of the outer contour in congruent patterns would provide a better fit for the tuning response properties of neurons in Area V1. These tuning properties, together with the limitations of the immature newborns' contrast sensitivity function, may act as a general perceptual filtering and amplification device, which would give congruent stimuli an attention-triggering advantage over noncongruent stimuli due to the regularity of the element position within the outer frame.

In addition to its relevance to the understanding of newborns' visual perceptual organization, the demonstration that the congruency property is capable of inducing visual preferences at birth in both facelike and nonfacelike stimuli is critical to the interpretation of the newborns' face preference phenomenon. Congruency is one of the perceptual properties contributing to the unique perceptual appearance of faces and facelike stimuli, together with top-heaviness. Overall, the results of the present study provide empirical support for this claim in two ways. First, Experiments 2, 3, and 4 demonstrated that the congruency property is necessary for the preference for facelike stimuli to manifest. Second, evidence from Experiment 5 showed that the copresence within a stimulus of the congruency and top-heavy perceptual properties produces a cumulative effect on newborns' visual preferences, which seems to be not dissimilar from the cumulative effect observed when congruency adds to facedness (Group 2 of Experiment 3).

More specifically, results of Experiments 1 and 2 show that the congruency property was capable of inducing a preferential response, which was not overcome by the preference for the facelike arrangement of the inner elements. Evidence from Experiment 3 demonstrates that the two stimulus properties have an additive effect on newborns' preferences. The key finding from this experiment is that, as already suggested by the results of Experiment 2, facedness and congruency have the same strength in triggering visual attention at birth and contribute to the same extent to the preference that newborns manifest for schematic facelike stimuli. This finding was further strengthened by the results of Experiment 4, where a congruent facelike pattern was preferred over a facelike pattern surrounded by a neutral, nonoriented contour. This evidence suggests that what is crucial in determining newborns' preference for schematic faces is not just the presence of a high-contrast contour surrounding the facelike arrangement of the inner elements but rather the presence of an oriented contour generating a congruent relation with those elements. Finally, the results of Experiment 5 extended the evidence provided by the previous experiments to top-heavy stimuli, which, albeit nonfacelike, share with faces the presence of more patterning in the upper part.

Together, these findings suggest that newborns' preference for faces does not depend exclusively, or entirely, on the facelike arrangement of the inner elements, which defines facedness. Rather, the results of the present study demonstrate that facedness is the emergent product of at least two different perceptual properties—top-heaviness and congruency—each of which is capable to elicit newborns' preferences when embedded in nonfacelike stimuli. In fact, the attentional biases for both these properties are thought to derive from basic constraints on the newborns' visual system (Acerra et al., 2002; Simion et al., 2002) that provide a comparable gain value for all visual patterns that possess certain structural/configural properties, irrespective of their resemblance to faces.

An aspect of the present findings, which merits discussion, relates to the lack of any significant result for the number of orienting responses variable. Although newborns looked significantly longer at the congruent nonfacelike stimulus in Experiment 1, at the congruent facelike stimuli in Experiments 3, 4, and 5, and at the congruent top-heavy stimulus in Experiment 5, they did not orient more frequently to these stimuli. This finding is not surprising, given that congruency is a configural property emerging from the spatial interrelations among the stimulus components that newborns may more easily appreciate when foveating the stimulus. Moreover, as proposed by other authors (e.g., Farroni, Menon, Rigato, & Johnson, 2007; Valenza et al., 1996), this finding may reflect the involvement of neural circuitry other than the subcortical visuomotor pathway that, according to the original version of the Johnson and Morton (1991) model of the development of face processing, would support the Conspec-like face detector responsible for newborns' face preference (Johnson, 2005).

Because stimuli in the periphery impinge on the temporal visual field that is known to feed predominantly into the subcortical visuomotor route, orienting responses are considered to likely reflect activity primarily within subcortical pathways (Simion, Valenza, Umiltà, & Dalla Barba, 1995, 1998).

To conclude, two final considerations should be raised concerning the interpretation of the data provided by the present study. First, we are aware that these data do not permit one to disentangle whether it is the same or rather two distinct factors that govern the preferences for facedness and the congruency property. In fact, the results of Experiments 2, 3, and 4 would be compatible with the interpretation that two distinct mechanisms may exist at birth, which work in parallel to drive newborns' attention. One of these would be a Conspec-like mechanism triggering newborns' attention toward stimuli displaying a facelike arrangement of the inner elements. The other mechanism could be nonspecific, responding to congruency and other general structural and configural properties of visual stimuli. According to this dual mechanism interpretation, the present results would show that the first mechanism is as powerful as the second in triggering newborns' attention and that the activation of the two mechanisms has an additive effect on newborns' preferences. However, because congruency is one of the configural/structural properties that characterize face stimuli, a more parsimonious explanation would be that newborns' visual behavior is governed by general rules that apply indifferently to faces and nonface stimuli. It has already been demonstrated that one of these rules concerns the up–down asymmetrical distribution of the features within the pattern (Simion et al., 2002). The present study provides evidence for a second rule concerning the spatial relation between the disposition of the inner features and the shape of the outer contour. Furthermore, examined more closely, the interpretation based on the existence of two separate mechanisms governing newborns' preferences for faces and nonface visual objects appears untenable. This is because, for Experiment 2, it would lead to predictions not confirmed by the actual results. Because the facelike stimulus used in Experiment 2 included both the facedness property and the general structural property of top-heaviness, this stimulus should have been preferred over the congruent nonfacelike stimulus in that it would activate both the face-specific and the general mechanisms. Instead, newborns' failure to manifest a preference for either stimulus renders more plausible the hypothesis of one single mechanism that would be sensitive to all the properties embedded in the stimulus, including top-heaviness and congruency, which together contribute to the overall facelike appearance of the stimulus.

A final aspect related to the interpretation of the present findings that should be acknowledged is the difficulty in unambiguously concluding that the attentional responses evoked by both the congruency and the top-heavy properties are of a nonspecific nature. Two different interpretations could be proposed as to why these two properties in combination trigger attention at birth. One possible interpretation is that the visual perceptual system has developed general attentional biases toward the congruency and top-heavy property, which in fact represent the nature's economical way of evolving a face detector. The alternative interpretation is that because facedness contains the congruency and top-heavy property as two of its defining characteristics, the innate face template matches better the nonface stimuli containing these two properties in comparison to other stimuli. Irrespective of the direction of this causal relationship, the demonstration provided by the present study of an additive effect of the two perceptual properties of congruency and top-heaviness on newborns' visual preferences appears crucial in the understanding of what makes a face to newborns.

References

Acerra, F., Burnod, I., & de Schonen, S. (2002). Modelling aspects of face processing in early infancy. *Developmental Science*, *5*, 98–117.

Antell, S. A., Caron, A. J., & Myers, R. S. (1985). Perception of relational invariants by newborns. *Developmental Psychology*, *21*, 942–948.

Banks, M., & Ginsburg, A. P. (1985). Early visual preferences: A review and new theoretical treatment. In H. W. Reese (Ed.), *Advances in child development and behavior* (Vol. 19, pp. 207–246). New York: Academic Press.

Banks, M., & Salapatek, P. (1981). Infant pattern vision: A new approach based on the contrast sensory function. *Journal of Experimental Child Psychology*, *31*, 1–45.

Cashon, C. H., & Cohen, L. (2004). Beyond U-shaped development in infants' processing of faces: An information-processing account. *Journal of Cognition and Development*, *5*, 59–80.

Cohen, L. B. (1972). Attention-getting and attention-holding process of infant visual preferences. *Child Development*, *43*, 869–879.

Cohen, L. B. (1973). A two process model of infant visual attention. *Merrill Palmer Quarterly*, *19*, 157–180.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General*, *115*, 107–117.

Farah, M. J., Rabinowitz, C., Quinn, G. E., & Liu, G. T. (2000). Early commitment of neural substrates for face recognition. *Cognitive Neuropsychology, 17,* 117 – 123.

Farroni, T., Menon, E., Rigato, S., & Johnson, M. H. (2007). The perception of facial expressions in newborns. *European Journal of Developmental Psychology, 4,* 2 – 13.

Farroni, T., Valenza, E., Simion, F., & Umiltà, C. (2000). Configural processing at birth: Evidence for perceptual organization. *Perception, 29,* 355 – 372.

Garner, W. R. (1974). *The processing of information and structure.* Hillsdale, NJ: Erlbaum.

Gauthier, I., & Tarr, M. J. (1997). Becoming a "greeble" expert: Exploring mechanisms for face recognition. *Vision Research, 37,* 1673 – 1682.

Gauthier, I., & Tarr, M. J. (2002). Unraveling mechanisms for expert object recognition: Bridging brain activity and behavior. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 431 – 446.

Geldart, S., Mondloch, C. J., Maurer, D., de Schonen, S., & Brent, H. P. (2002). The effect of early visual deprivation on the development of face processing. *Developmental Science, 5,* 490 – 501.

Johnson, M. H. (2000). Functional brain development in infants: Elements of an interactive specialization framework. *Child Development, 71,* 75 – 81.

Johnson, M. H. (2005). Subcortical face processing. *Nature, 6,* 766 – 774.

Johnson, M. H., & Morton, J. (1991). *Biology and cognitive development. The case of face recognition.* Oxford, England: Basil Blackwell.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17,* 4302 – 4311.

Kimchi, R. (1992). Primacy of wholistic processing and global/local paradigm: A critical review. *Psychological Bulletin, 112,* 24 – 38.

Le Grand, R., Mondloch, C. J., Maurer, D., & Brent, H. P. (2001). Early visual experience and face processing. *Nature, 410,* 890.

Macchi Cassia, V., Simion, F., Milani, I., & Umiltà, C. (2002). Dominance of global visual properties at birth. *Journal of Experimental Psychology: General, 4,* 398 – 411.

Macchi Cassia, V., Turati, C., & Simion, F. (2004). Can a nonspecific bias toward top-heavy patterns explain newborns' face preference? *Psychological Science, 15,* 379 – 383.

Mondloch, C. J., Lewis, T. L., Budreau, D. R., Maurer, D., Dannemiller, J. L., Stephens, B. R., et al. (1999). Face perception during early infancy. *Psychological Science, 10,* 419 – 422.

Palmer, S. E. (1991). Goodness, Gestalt, groups, and Garner: Local symmetry subgroups as a theory of figural goodness. In G. R. Lockhead & J. R. Pomerantz (Eds.), *The perception of structure: Essays in honor of Wendell R. Garner* (pp. 23 – 39). Washington, DC: American Psychological Association.

Peterson, M. A., & Rhodes, G. (2003). *Perception of faces, objects, and scenes. Analytical and holistic processes.* New York: Oxford University Press.

Rubin, N., Nakayama, K., & Shapley, R. (1996). Enhanced perception of illusory contours in the lower versus upper visual hemifield. *Science, 271,* 651 – 653.

Schwarzer, G., Zauner, N., & Jovanovic, B. (2007). Evidence of a shift from featural to configural face processing in infancy. *Developmental Science, 10,* 452 – 463.

Sergent, J., Ohta, S., & MacDonald, B. (1992). Functional neuroanatomy of face and object processing. *Brain, 115,* 15 – 36.

Simion, F., Macchi Cassia, V., Turati, C., & Valenza, E. (2001). The origins of face perception: Specific vs. non-specific mechanisms. *Infant and Child Development, 10,* 59 – 65.

Simion, F., Macchi Cassia, V., Turati, C., & Valenza, E. (2003). Non-specific perceptual biases at the origins of face processing. In O. Pascalis & A. Slater (Eds.), *The development of face processing in infancy and early childhood: Current perspectives* (pp. 13 – 25). New York: Nova Science.

Simion, F., Valenza, E., Macchi Cassia, V., Turati, C., & Umiltà, C. (2002). Newborns' preference for up-down asymmetrical configurations. *Developmental Science, 5,* 427 – 434.

Simion, F., Valenza, E., Umiltà, C., & Dalla Barba, B. (1995). Inhibition of return in newborns is temporo-nasal asymmetrical. *Infant Behavior and Development, 18,* 189 – 194.

Simion, F., Valenza, E., Umiltà, C., & Dalla Barba, B. (1998). Preferential orientino to faces in newborns: A temporal-nasal asymmetry. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1399 – 1405.

Skrandies, W. (1987). The upper and lower visual field of man. Electrophysiological and functional differences. In H. Autrum (Ed.), *Progress in sensory physiology* (Vol. 8, pp. 162 – 189) Germany: Springer-Verlag.

Slater, A. (1995). Visual perception and memory at birth. In C. Rovee-Collier & L. P. Lipsitt (Eds.), *Advances in infancy research* (Vol. 9, pp. 107 – 162). Norwood, NJ: Ablex.

Slater, A., Earle, D. C., Morison, V., & Rose, D. (1985). Pattern preferences at birth and their interaction with habituation-induced novelty preferences. *Journal of Experimental Child Psychology, 39,* 37 – 54.

Slater, A., & Sykes, M. (1977). Newborn infants' responses to square-wave gratings. *Child Development, 48,* 545 – 553.

Turati, C. (2004). Why faces are not special to newborns: An alternative account of the face preference. *Current Direction in Psychological Science, 13,* 5 – 8.

Turati, C., Simion, F., Milani, I., & Umiltà, C. (2002). Newborns preference for faces: What is crucial? *Developmental Psychology, 6,* 875 – 882.

Valenza, E., & Bulf, H. (2007). The role of kinetic information in newborns' perception of illusory contours. *Developmental Science, 10,* 492 – 501.

Valenza, E., Simion, F., Macchi Cassia, V., & Umiltà, C. (1996). Face preference at birth. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 892 – 903.

Child Development, July/August 2008, Volume 79, Number 4, Pages 821–835

Grownups Are Not Afraid of Scary Stuff, but Kids Are: Young Children's and Adults' Reasoning About Children's, Infants', and Adults' Fears

Liat Sayfan and Kristin Hansen Lagattuta
University of California, Davis

Three-, 5-, and 7-year-olds and adults ($N = 64$) listened to stories depicting 2 protagonists of different ages (infant and child or child and grownup) that encounter an entity that looks like a real (e.g., a snake) or an imaginary (e.g., a ghost) fear-inducing creature. Participants predicted and explained each protagonist's intensity of fear. Results showed significant age-related increases in knowledge that infants and adults would experience less intense fears than young children and that people's fears are causally linked to their cognitive mental states. Across age, stories involving imaginary beings elicited more frequent mental explanations for fear than stories about real creatures. Results are discussed in relation to children's developing awareness of the mind as mediating between situations and emotions.

Fear is one of the four basic emotions children experience, express, and identify in others from very early in life (Ekman, Sorenson, & Friesen, 1969). By using parent-report and child-report surveys, studies have found that 75%–95% of normal children experience an average of three to five specific fears of mild to moderate intensity (Gullone, 2000; Muris, Merckelbach, Gadet, & Moulaert, 2000). By the age of 3 years, most children exhibit fears of bodily injury, pain, natural phenomena (e.g., earthquakes), the dark, and animals (e.g., snakes). During the preschool years, children also begin to fear imaginary and supernatural creatures, including witches, monsters, and ghosts, with these fears of the imaginary and supernatural typically diminishing substantially by school age (Muris et al., 1997; Muris & Merckelbach, 2000).

Although we have ample information about the situations, objects, and creatures that young children fear, researchers have not systematically examined developmental changes in children's knowledge about the sources of fears as well as their awareness that people of different ages may experience different intensities of fear in the *same exact* situation. The current research addresses this missing piece by exploring 3- to 7-year-olds' knowledge about age differences in fears of real and imaginary creatures. In doing so, the current study provides a window into young children's awareness that people's emotions are not just elicited by objective situations but are also strongly shaped by the person's knowledge (e.g., knowledge about what is real vs. imaginary, knowledge vs. ignorance about an entity's capability to harm), characteristics (e.g., size and gender), and preferences (i.e., likes and dislikes).

Indeed, fear is an ideal context for examining children's early concepts about individual differences in emotions, especially how diversity in emotions may stem from differences in people's mental states. That is, fears are commonly elicited by thinking about the potential for harm, or in the case of fears of supernatural creatures, imagining something that does not exist. For this reason, fears commonly change with age and experience, as children increase their knowledge base as well as decrease their participation in fantasy play (Sarafino, 1986; Wenar, 1990). Moreover, it is likely that young children frequently encounter people with varying levels of fear in their everyday experiences, as when a mother chides, "You shouldn't be afraid—ghosts aren't real!" The current study examines whether 3- to 7-year-olds recognize that children typically experience more intense fears, especially fears of imaginary creatures, compared to adults (who know more) or young infants (who are naïve). Alternatively, young children may assume that fear depends just on the situation, rather than people's knowledge or attributes. If so, they may predict that all people will experience the same level of fear.

Prior research on children's emotion understanding reveals that most preschoolers have strong

---

We thank Colby Chlebowski and Lakeshia Mcgee for their assistance with this project. We also want to thank Helit Sayfan Altman for creating the wonderful illustrations. Gratitude also goes to Christi Bamford for her comments on initial drafts of this article.

Correspondence concerning this article should be addressed to Liat Sayfan, Department of Psychology, Center for Mind and Brain, University of California, 1 Shields Avenue, Davis, CA 95616. Electronic mail may be sent to lsayfan@ucdavis.edu.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7904-0002

knowledge about situational causes of emotions in that they can easily pair prototypic situations with common emotions (Harris, 1989). However, when a person's emotion cannot be easily explained by external events alone (e.g., feeling sad at a birthday party), young children often fail to consider the person's attributes, attitudes, or experiences (Denham, 1986; Gnepp, 1983; Gnepp & Gould, 1985). More recent research, however, reveals that young children have more than just a situational or script-like understanding of emotions (see Thompson & Lagattuta, 2006, for a review). Notably, even 4- to 6-year-olds can explain people's emotions as caused by their thoughts, memories, and anticipations (Lagattuta, 2007; Lagattuta & Wellman, 2001; Lagattuta, Wellman, & Flavell, 1997). Moreover, Rieffe, Meerum Terwogt, and Cowan (2005) report that 4-year-olds exhibit stronger awareness about mental causes of emotions in fear-inducing situations compared to happy-, anger-, or sad-arousing situations. Thus, although previous studies have been mixed, there is mounting evidence that young children know that both attributes of situations and people's minds determine emotions.

Finally, clues as to when young children may understand age-related differences in fear come from studies on children's knowledge that people of different age groups have different levels of factual knowledge (see Miller, 2000, for a review). That is, because fear is largely based on beliefs about potential harm from an object, entity, or situation (Oatley & Jenkins, 1996), different knowledge levels should translate into different fear experiences. Research has shown that although 4-year-olds appreciate that adults know more than children and children know more than babies, they have difficulty applying this awareness of knowledge diversity to make predictions about how people of different age groups would interpret partly occluded objects (Taylor, Cartwright, & Bowden, 1991). Further studies have shown that 4- to 8-year-olds exhibit the most difficulty assessing infants' knowledge and typically err on the side of overestimating what infants can know or do (Dowker, Hart, Heal, Phillips, & Wilson, 1994; Miller, Hardin, & Montgomery, 2003; Montgomery, 1993). More specific to the current research, in an unpublished dissertation, Phelps (1999) found that whereas 3-year-olds thought that adults are fearful of both real and imaginary entities, 4- and 5-year-olds asserted that adults are generally not afraid of most things. Nevertheless, all age groups maintained that babies would be afraid of monsters.

In summary, then, the current research examines developmental changes in young children's knowledge that people of different ages may have different fear reactions to identical situations. We presented 3- to 7-year-old children with vignettes featuring two people of different ages (a child and an infant or a child and a grownup) that encounter a stimulus that looks like a real (bee, spider, snake, and alligator) or an imaginary (witch, monster, ghost, and dragon) creature that preschool children commonly fear. We asked children to predict and explain the emotions of the story characters, including the intensity of each person's fear reaction (if afraid), using a 4-point pictorial intensity scale. Fear intensity was used because it enabled investigation into whether children believe the emotions of different age groups differ in *type* (afraid vs. unafraid) or whether they differ in degree of *intensity* (e.g., less afraid vs. more afraid).

Because prior research has been mixed as to whether young preschoolers understand individual or age-related differences in people's knowledge or emotions, for the last four story trials, we explicitly provided children with a direct comparison of the two people's contrasting emotional reactions (e.g., "[The child] is more afraid than the grownup right now") and asked them to provide an explanation. We reasoned that young children would provide more sophisticated person-oriented (particularly mention of cognitive mental states) explanations for emotions when they were explicitly prompted about divergent emotions (the *explain-different emotion* trials) versus when justifying their own emotion predictions.

Finally, because children's persistent fear of imaginary creatures has raised the question as to their ability to discern real entities from imaginary ones (Rosengren, Johnson, & Harris, 2000), at the end of the story-card procedures, we interviewed children about their own fears of the stimuli used in the vignettes as well as their understanding about the reality status of the imaginary creatures. College undergraduates were tested using the same procedures to measure adult reasoning and beliefs.

Method

Participants

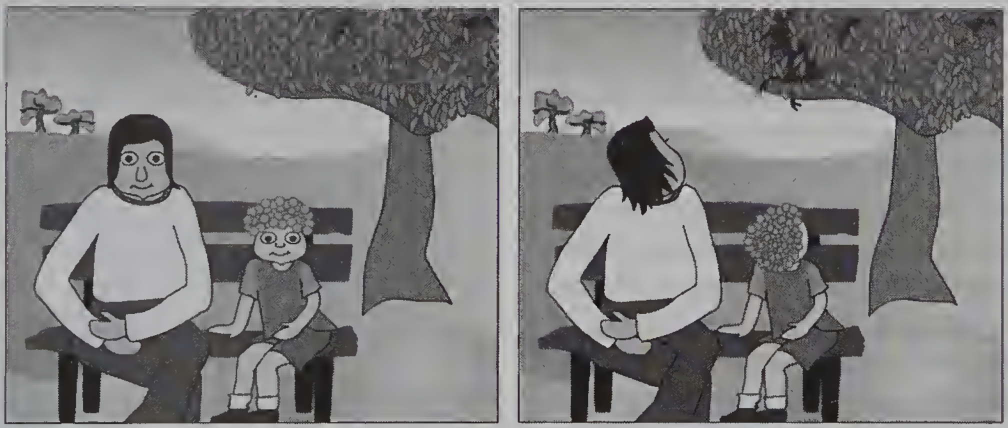
Sixty-four individuals in four age groups participated: sixteen 3-year-olds (age: $M = 3$ years 8 months, range = 3 years 4 months to 4 years 1 month), sixteen 5-year-olds ($M = 5$ years 4 months, range = 5 years 0 month to 5 years 10 months), sixteen 7-year-olds ($M = 7$ years 5 months, range = 7 years 0 month to 7 years 10 months), and sixteen young adults ($M = 20$ years 9 months, range = 18 years 8 months to 26 years

7 months), with equal numbers of males and females in each age group. Five additional 3-year-olds participated but never completed the interview due to inability or unwillingness to respond. The children were recruited from local schools and child-care centers in an ethnically diverse middle-class university town. Adults were recruited using the university introductory psychology subject pool. Of the adults, 50% were Caucasian, 25% Asian, 12.5% Hispanic, and 12.5% Other. Data were not collected on children's ethnicity, but they were predominantly Caucasian (approximately 75%).

Materials

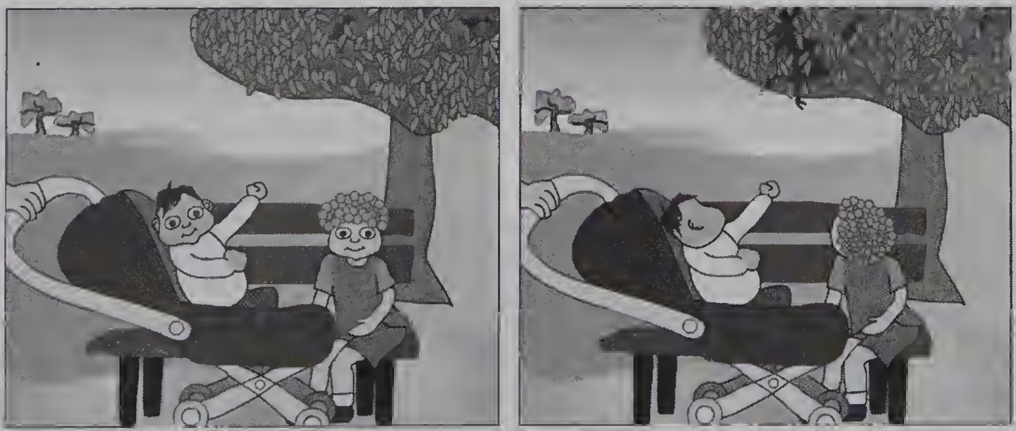
Participants listened to eight short stories accompanied by pictures illustrating each scene. Each story involved two protagonists: (a) a preschool boy or a preschool girl (the *child character*) and (b) either an androgynous grownup or an androgynous baby. In each story, the two protagonists engaged in a neutral behavior (e.g., sitting on the sand or standing near the stairs). Suddenly, they notice a potentially fear-inducing stimulus. The fear-inducing stimulus was described as either looking like a *real* ("It looks like a bee/a spider/an alligator/a snake") or an *imaginary* creature ("It looks like a ghost/a witch/a dragon/a monster"). The phrase *looks like* was included for both real and imaginary trials to make the wording linguistically parallel. Each stimulus was drawn so that it looked like a potentially scary thing, but it was unclear whether or not it actually was a scary thing (e.g., the snake could be interpreted as a worm or the ghost as a hanging piece of cloth). The wording of each story remained consistent across all stories with the exception of the story location, the characters involved, and the type of creature. Figure 1 shows the two versions of an imaginary creature story (the witch story). The child–grownup version is presented at the top of the figure and the child–baby version is presented at the bottom.

Each participant heard eight stories (about a spider, a snake, a bee, an alligator, a monster, a witch, a ghost,



Child-Grownup Version: "This is a story about a boy named Joe. One day Joe and his grownup friend are at the park. They sit together on a bench"

"Suddenly, they see something in the tree. It looks like a witch."



Child-Baby Version: "This is a story about a boy named Joe. One day Joe and his baby friend are at the park. They sit together on a bench"

"Suddenly, they see something in the tree. It looks like a witch."

Figure 1. Example of an imaginary creature story (the witch) with child–grownup pair (top) and child–baby pair (bottom).

and a dragon) in one of eight predetermined order sets. Each order set included two stories about a real entity with a child–grownup pair of protagonists, two stories about a real entity with a child–baby pair, two stories about an imaginary entity with a child–baby pair, and two stories about an imaginary entity with a child–grownup pair. Two of the real-entity stories had a girl as the child character (spider and snake stories) and two had a boy (bee and alligator stories). Similarly, two of the imaginary-entity stories were about a girl (monster and ghost stories) and two were about a boy (dragon and witch stories). To make sure that children interpreted these scenarios as events that could actually happen, the first story trial always involved a real creature. Additionally, because both the bee and the spider are similar stimuli, we included one in the first four story trials (Block 1) and one in the four last story trials (Block 2). In the same way, we separated the alligator and snake stories (both involve reptiles). The question order (whether participants predicted the child's vs. the grownup's or infant's reaction first) was counterbalanced across story trials.

Procedure

Each participant was interviewed individually in a quiet room in his or her school or in the laboratory. Before the interview began, the experimenter said, "You're going to hear a few short stories. I'm going to ask you questions about the people in the stories. There is no 'right' or 'wrong' answer; whatever you tell me is fine. I just want to know what you think." They were then pretrained on how to interpret and use the 4-point fear intensity scale (see Figure 2 for the scale illustration). Participants then heard eight short stories involving a potentially fear-inducing stimulus. Following each story, the experimenter asked the participant to predict each of the character's emotions (e.g., "Do you think the baby is afraid or not afraid right now?"), to specify the intensity of each of their fears (if predicted) using the 4-point scale (e.g., "How afraid is the baby?"), and to explain the cause of each of their emotions (e.g., "Why is the baby [not] afraid right now?"). To help participants clarify and expand their explanations, the experimenter paraphrased their responses and encouraged them to provide additional information. If children answered "I don't know," they were reminded that there were no right or wrong answers and encouraged to take a guess. When it was clear that the participant had no further explanations (after being asked three times), the experimenter proceeded to the next trial.

During the last four story trials, after participants provided their own emotion predictions and explanations, the experimenter explicitly told and showed them that the two characters were experiencing different levels of fear and she asked them to explain the reason for this difference (the explain-different emotion trials). Specifically, participants were asked: "Why do you think the baby/grownup is less/more afraid than [Joe] right now?" (the use of the terms *more* vs. *less* was counterbalanced across trials). Because prior research has shown that 3-year-olds may better comprehend the meaning of the linguistic term *more* versus *less* (even though they know that both are comparative terms; Palermo, 1974; Wannemacher & Ryan, 1978), this question was always accompanied by a story card that clearly juxtaposed the different intensities of fear experienced by the two protagonists. If children showed any confusion about how to interpret the question, it was rephrased with the other comparative term (e.g., "Why is Joe more afraid than the baby" instead of "Why is the baby less afraid than Joe?"). Figure 3 shows the picture cards used for the explain-different emotion trials for the child–grownup (top) and child–baby (bottom) versions of the witch story.

After participants responded to the eight story trials, the experimenter asked them whether or not they, themselves, feared the creatures presented in the stories (e.g., "Are you afraid or not afraid of spiders?"). If they said that they were afraid, the experimenter asked them to indicate the intensity of this

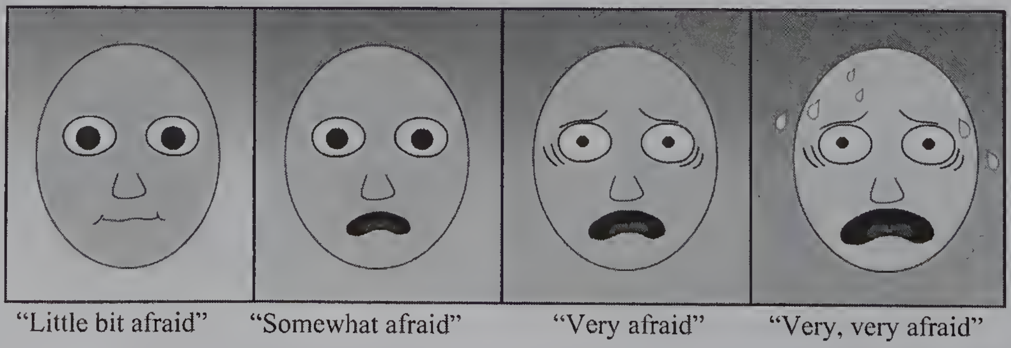
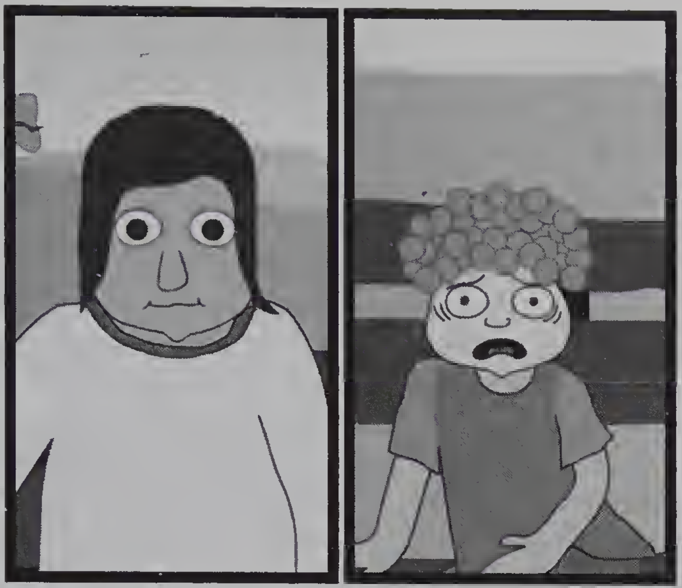
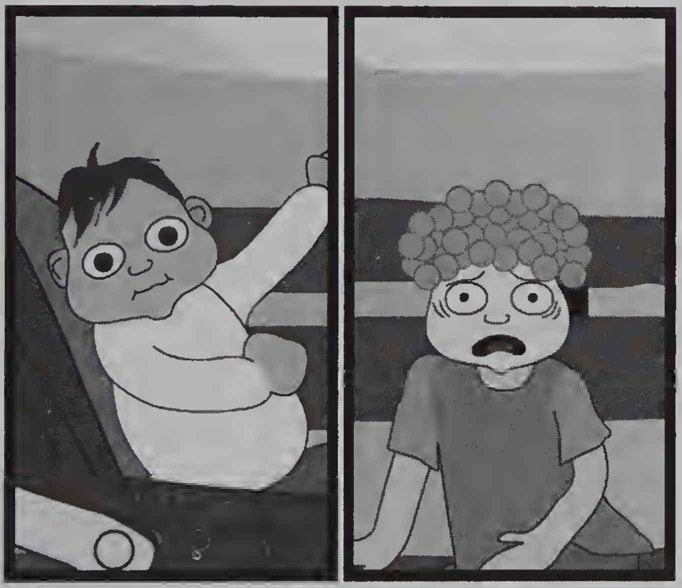


Figure 2. The fear intensity scale.



Child-Grownup Version: "Now, let's see how Joe and his friend are really feeling. Look! The grownup is less afraid than Joe."



Child-Baby Version: "Now, let's see how Joe and his friend are really feeling. Look! Joe is more afraid than the baby."

Figure 3. Examples of the explain-different emotion ending for the witch story with child–grownup pair (top) and child–baby pair (bottom).

fear on the pictorial scale (i.e., "Show me how afraid you are"). Finally, children were asked to judge whether each stimulus really existed (e.g., "Are there any real witches in the world?"). To verify that the reality question was interpreted correctly for the imaginary creatures, whenever participants validated the existence of an imaginary creature, the experimenter asked them to explain how they view the entity (e.g., "When you say witches are real, what do you think a witch is?"). This enabled us to count only endorsements that fit supernatural criteria. For example, if a participant said, "Yes, there are real dragons," but then went on to mention the large reptiles known as komodo dragons, this would not be counted as an endorsement of imaginary dragons (for witches, they had to mention something about being able to do magic or cast spells; for ghosts, it had to be an incarnation of a dead person; and for monsters, it could not simply be a real animal with a deformity).

Each session was tape-recorded and transcribed verbatim. After completion of the session, child participants received a small reward and adult participants were given course credit for their participation.

Coding of Fear Explanations

Participants' explanations for the story characters' fear reactions were coded into two main categories: *stimulus oriented* and *person oriented* (see Table 1 for examples). Stimulus-oriented explanations focused on the features of the creature itself, including its identity, size, and capability to harm. In contrast, person-oriented explanations explained emotions in relation to features or qualities of the story character. There were three subtypes of person-oriented explanations. *Attribute explanations* included mention of the protagonist's external characteristics (e.g., size), gender, or age. *Preference explanations* attributed the cause of emotions to the protagonist's desires, likes, or dislikes and included verbs such as like, love, hate, dislike, or want. Finally, *mind explanations* focused on the protagonist's cognitive mental states and included mental verbs such as *know, believe, think*, or *realize*. Because participants often provided multiple emotion explanations for a single character, their complete explanation could be coded into one or into multiple categories. For example, the explanation, "[He's afraid because] the alligator might come closer. He doesn't want the grownup to smash it down. The alligator might splash water on him" would be coded as both stimulus oriented and person oriented (preference). This is because the child explained fear in relation to the alligator potentially coming closer and splashing water (stimulus oriented) as well as in relation to what the character does not want to happen (person oriented, preference).

The principal investigator and two undergraduate research assistants collaboratively coded 25% of the transcripts. After this training session, the research assistants, who were blinded to the study hypotheses and to what were considered more sophisticated explanations, independently coded the remaining 75% of the transcripts into the different categories of explanations using explicit criteria. All disagreements were resolved by discussion. Kappa coefficients for the different explanation categories, including all subcategories, were high, ranging between .91 and .96.

Table 1

Examples of Children's Explanations for Why Characters Are Afraid or Not Afraid by Age Group and Explanation Type

	Age group			
Explanation type	3-year-olds	5-year-olds	7-year-olds	Adults
Stimulus oriented	Because *the snake is nice.*	The *witch is gonna take him* away and never come back.	*It's* just might be *a bug*.	*Witches don't attack* in daylight.
	Because *it's not getting out* of the water.		It's just a *grizzly bear.*	*Alligators can kill* you.
Person oriented: Mind	*She thinks* ghosts are scary.	He's scared because *he thinks* the dragon will see him and blow fire at him and he's gonna burn up.[a]	He's at the age when *he believes* in those things.	The baby *doesn't know* what is going on.
	Because *she thought* it's a monster.		*He thinks* it is just a shadow.	*Pete knows* its just a fish.[a]
			He thinks the witch is gonna get him and put him in stew.[a]	She *doesn't know* that spiders don't hurt people.[a]
Person oriented: Preferences	The baby's afraid because it *doesn't like* spiders.	*He doesn't want* the alligator to eat him.[a]	The grownup *likes* snakes.	*He doesn't want* to be around this thing.
	She *doesn't want* to have a dragon there.		*He doesn't want* the witch to hurt him.[a]	
Person oriented: Attributes	Because *he's just a tiny, tiny baby.*	Grownups aren't afraid. *They're bigger* than little kids.	*She can't run* as fast.	Because *little girls* get scared of spiders.
			Babies get scared when they see stuff that's big.[a]	*She's* only *a child.*

Note. Coders made decisions about the explanations' classifications based on the italicized key phrases.

[a] These explanations were also coded as stimulus oriented because they describe features or capabilities of the entity as well.

Results

Several approaches were taken to investigate developmental changes in children's knowledge that people of different age groups would experience different emotions in the same situation. First, we focus on the number of trials that children and adults predicted that the emotions of the two story characters would differ by *type* (one is afraid and the other is not afraid). Next, we examine the number of trials that children and adults predicted that the characters' fear reactions would differ in *intensity* (one is more or less afraid than the other). Final prediction analyses examine whether predictions about the intensity of characters' fear reactions varied by the age of the character (baby vs. child vs. grownup) or the reality status of the stimulus (real vs. imaginary). Prediction analyses are then supplemented by an examination of children's and adults' *explanations* for the causes of characters' emotions.

Because preliminary analyses revealed no significant main effects or interactions for story order or gender of child story character, these factors were not considered further (see individual sections subsequently for how data were scored). Participant gender was also removed from the primary analyses because it had no significant main effects or interactions on predictions or explanations aside from one exception. There was a significant main effect for participant gender for fear intensity predictions, $F(1, 56) = 5.78$, $p < .05$, with males ($M = 1.34$, $SE = 0.10$) attributing more intense fear to story characters than females ($M = 1.01$, $SE = 0.11$). Finally, 3-year-olds did not show greater confusion when the explain-different-emotion question was phrased as "less" versus "more" afraid. They answered "I don't know" as their first response for 22% of the trials, regardless of comparative term.

Except when otherwise noted, data were analyzed using repeated measures analysis of variance (ANOVA). Alpha was set at .05, and thus, all results reported as significant are $p < .5$ or better. Significant main effects were followed up with Tukey's honestly significant difference (HSD) tests to evaluate pairwise comparisons, with only significant pairwise comparisons among means reported. Simple effects tests were used to examine significant interactions. Partial eta squares, computed using SPSS, are provided as a measure of effect size for all significant main effects and interactions.

Analyses of Fear Predictions

Predicting fear versus nonfear reactions. First, we examined the frequency that children and adults predicted that two characters' emotions in the same situation would vary by type (dichotomously): afraid versus not afraid. Participants were given a score of 1 for every trial that they predicted that the pair of characters (child – baby or child – grownup) would have different reactions (one character is afraid and the other is not afraid). A score of 0 was given when participants predicted both characters would have the same reaction (i.e., both "afraid" or both "not afraid"). All scores were averaged across trials.

A 4 (age: 3-year-olds, 5-year-olds, 7-year-olds, and adults) × 2 (entity type: real vs. imaginary) × 2 (pair: child – baby vs. child – grownup) repeated measures ANOVA for predict-different-emotion scores resulted in a main effect for entity type, $F(1, 60) = 6.53$, $\eta_p^2 = .10$, with all age groups predicting dissimilar fear reactions significantly more often in situations that involved imaginary entities ($M = 0.54$, $SE = 0.04$) than in situations involving real entities ($M = 0.45$, $SE = 0.04$). Separate post hoc t tests, however, showed that all participants, even the adults, were at chance level (50%) when attributing differential fear versus nonfear reactions to characters, $ts(15) < 1.27$, $ps > .22$.

Predicting differential intensity of fear. Next, we examined whether children and adults predicted that two people in the same situation would vary in the *intensity* of their fear reactions (more vs. less afraid). Participants received a score of 1 for every trial that they predicted that each character in the pair (child – baby or child – grownup) would experience a different level of fear (one character is *more* afraid than the other or one character is afraid and the other is not). A score of 0 was given when participants predicted both characters would have the same *intensity* of fear reaction. Again, these scores were averaged across trials.

A 4 (age: 3-year-olds, 5-year-olds, 7-year-olds, and adults) × 2 (entity type: real vs. imaginary) × 2 (pair: child – baby vs. child – grownup) repeated measures ANOVA yielded main effects for age, $F(1, 60) = 3.5$, $\eta_p^2 = .15$, and for entity type, $F(1, 60) = 5.2$, $\eta_p^2 = .08$. The frequency that participants predicted that the two characters' fear reactions would differ in intensity significantly increased with age (3-year-olds: $M = 0.50$, 5-year-olds: $M = 0.63$, 7-year-olds: $M = 0.70$, and adults: $M = 0.80$, and all SEs $= 0.05$). Post hoc Tukey's HSD comparisons showed that 3-year-olds predicted different fear intensities for characters significantly less often than 7-year-olds or adults and that 5-year-olds also made such predictions significantly less often than adults. Additionally, participants of all age group predicted dissimilar fear reactions significantly more often for stories about imaginary entities ($M = 0.70$, $SE = 0.04$) versus real entities ($M = 0.61$,

$SE = 0.04$). Separate t tests showed that 3-year-olds predicted different emotions at above chance levels for the stories about imaginary entities, $ts(15) > 3.1$; however, they were at chance when predicting different emotions for characters encountering real entities, $ts(15) < 2.09$. All other age groups predicted different emotions at greater than chance levels across all entity types, $ts(15) > 2.6$.

Fear intensity attributed to characters across story trials. Next, we explored developmental trends in predicting different intensities of fear (ranging from 0 = *no fear* to 4 = *intense fear*) for people in relation to their age (infants, children, and grownups) and the reality status of the stimuli (real vs. imaginary). Because the child character appeared in every trial, child character scores were averaged across the four trials for each entity type (real or imaginary). Scores for companion characters (baby and grownup) were averaged across the two trials for each entity type that each character appeared. Table 2 and Figure 4 display the primary results. A 4 (age: 3-, 5-, and 7-year-olds and adults) × 3 (character: infant, child, and grownup) × 2 (entity type: real vs. imaginary) repeated measures ANOVA for fear intensity predictions revealed main effects for character, $F(2, 120) = 24.39$, $\eta_p^2 = .29$, and for entity type, $F(1, 60) = 14.88$, $\eta_p^2 = .20$, a significant Character × Entity Type interaction, $F(2, 120) = 7.39$, $\eta_p^2 = .11$, and a significant Age Group × Character interaction, $F(6, 120) = 7.49$, $\eta_p^2 = .27$; all qualified by a three-way interaction between age group, character, and entity type, $F(6, 120) = 3.02$, $\eta_p^2 = .13$.

To clarify the three-way interaction, we performed separate 3 (character: infant, child, and grownup) × 2 (entity type: real vs. imaginary creatures) repeated measures ANOVAs for each age group. For the 3- and 5-year-old group, this resulted only in a main effect for character, $F(2, 30) = 8.12$, $\eta_p^2 = .35$, and $F(2, 30) = 10.85$, $\eta_p^2 = .42$, respectively. For the 7-year-olds and the adults, there was a significant Character × Entity Type interaction, $F(2, 30) = 9.50$, $\eta_p^2 = .39$, and $F(2, 30) = 7.33$, $\eta_p^2 = .33$, respectively, qualifying a main effect for entity type for the 7-year-olds, $F(1, 15) = 9.83$, $\eta_p^2 = .40$, and a main effect for character for the adults, $F(1, 15) = 32.51$, $\eta_p^2 = .68$.

Table 2

Participants' Predictions of Fear Intensity by Age Group and Story Character

	Story character		
Age group	Infant	Child	Grownup
3-year-olds	1.58 (1.30)	1.02 (1.05)	0.45 (0.74)
5-year-olds	2.19 (1.60)	1.43 (1.00)	0.73 (0.94)
7-year-olds	1.29 (1.04)	1.27 (0.87)	0.83 (0.77)
Adults	0.88 (0.97)	2.12 (0.71)	0.89 (0.92)

Note. The fear scale ranged from 1 (*little bit afraid*) to 4 (*very very afraid*). The means presented here also include the prediction of *not afraid* coded as 0. The same pattern of results emerged when we excluded "not afraid" predictions and considered only fear predictions. Standard deviations are presented in parentheses.

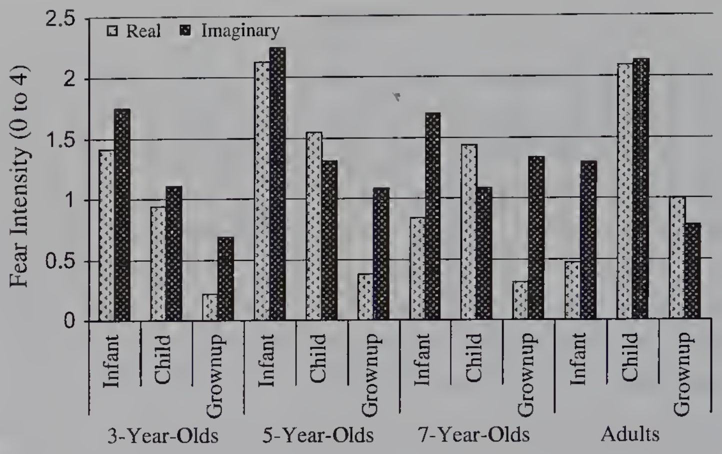


Figure 4. Participants' fear predictions by age group, story character, and entity type.

As shown in Table 2, 3- and 5-year-olds attributed more intense fears to the infant characters followed by the child characters and the least intense fears to the grownup characters. Similarly, 7-year-olds attributed the least intense fears to the grownup characters; however, they predicted equivalent fear intensities for the child and infant characters. Finally, adults judged that child characters would experience the most intense fears and infants and grownups would experience the least intense fears.

Figure 4 shows data on participants' fear predictions by entity type. As shown in Figure 4, within the stories about *real* creatures, both 7-year-olds and adults predicted the most intense fears for children ($M = 1.44$, $SE = 0.24$ and $M = 2.10$, $SE = 0.24$, respectively), followed by both infants ($M = 0.84$, $SE = 0.29$ and $M = 0.47$, $SE = 0.29$, respectively) and grownups ($M = 0.31$, $SE = 0.17$ and $M = 1.00$, $SE = 0.17$, respectively). Within the stories about *imaginary* creatures, 7-year-olds predicted that children would be less afraid ($M = 1.09$, $SE = 0.22$) than both grownups ($M = 1.34$, $SE = 0.22$) and infants ($M = 1.70$, $SE = 0.33$). However, adults predicted that children would be the most afraid ($M = 2.14$, $SE = 0.22$) compared to both grownups ($M = 0.78$, $SE = 0.22$) and infants ($M = 1.30$, $SE = 0.33$). Finally, simple effects tests revealed that 7-year-olds predicted that both infants and grownups would be more afraid of imaginary things ($M = 1.70$, $SE = 0.33$ and $M = 1.34$, $SE = 0.22$, respectively) than real things ($M = 0.84$, $SE = 0.29$ and $M = 0.31$, $SE = 0.17$, respectively) and that children will be equally afraid of both real

($M = 1.44$, $SE = 0.24$) and imaginary ($M = 1.09$, $SE = 0.22$). On the other hand, adults predicted that infants would be more afraid of imaginary things ($M = 1.30$, $SE = 0.33$) than real things ($M = 0.47$, $SE = 0.29$), whereas children and grownups would be equally afraid of both real ($M = 2.10$, $SE = 0.24$ and $M = 1.00$, $SE = 0.17$, respectively) and imaginary creatures ($M = 2.14$, $SE = 0.22$ and $M = 0.78$, $SE = 0.22$, respectively).

Emotion Explanations

As described in the Method section, participants' emotion explanations were coded into several categories (Table 1). Participants were given a score of 1 for every trial that they provided a particular type of explanation for each character's emotion (i.e., person oriented or stimulus oriented, including the different subtypes of person-oriented explanations). Explanation scores were averaged across the number of trials in which each character appeared for each entity type (i.e., across the four imaginary and four real trials for child characters and across the two imaginary and two real trials for grownup and infant characters). Additionally, we averaged the number of trials participants offered both types of explanations at the same time (i.e., a combined stimulus- and person-oriented explanation).

Table 3 displays these data, converted to percentage of trials for ease of interpretation. A 4 (age group: 3-, 5-, and 7-year-olds and adults) × 3 (explanation type: stimulus oriented, person oriented, and both) repeated measures ANOVA revealed main effects for explanation type, $F(2, 120) = 11.71$, $\eta_p^2 = .16$, and age group, $F(3, 60) = 4.27$, $\eta_p^2 = .18$, and a significant Age Group × Explanation Type interaction, $F(6, 120) = 4.40$, $\eta_p^2 = .18$. As shown in Table 3, exclusive stimulus-oriented explanations significantly decreased with age as combined explanations referring to both the stimulus and the person significantly increased. There were no differences between the age groups for the frequency of exclusive person-oriented explanations.

To further elucidate the developmental progression of children's emotion explanations, we examined age-related changes in the provision of the three subcategories of person-oriented explanations. Three separate repeated measures ANOVAs on mind explanations, preference explanations, and attribute explanations were conducted, with age group as the between-subject factor and character and entity type as the within-subjects factors. Table 4 shows these data. For the mind explanations, we found main effects for age group, $F(3, 60) = 18.99$, $\eta_p^2 = .49$, and for entity type, $F(1, 60) = 20.86$, $\eta_p^2 = .26$. Specifically, 5- and 7-year-olds and adults provided mind explanations significantly more often than 3-year-olds. Additionally, all age groups explained emotions in relation to characters' cognitive mental states significantly more often in situations involving imaginary versus real things. For the preference explanations, there was only a main effect for age group, $F(3, 60) =$

Table 3
Stimulus- and Person-Oriented Explanations by Age Group and Entity Type[a]

	Age group			
Entity type	3-year-olds	5-year-olds	7-year-olds	Adults
Exclusive stimulus-oriented explanations				
Real[b]	43	27	31	25
Imaginary[b]	35	27	27	13
Total[c]	39	27	29	19
Exclusive person-oriented explanations				
Real[b]	28	27	28	16
Imaginary[b]	22	11	14	20
Total[c]	25	19	21	18
Combined stimulus- and person-oriented explanations				
Real[b]	19	41	33	56
Imaginary[b]	23	55	47	64
Total[c]	21	48	40	60

[a] All numbers are percentages of trials. Note that percentages for each group do not add up to 100% because some explanations did not fit any of the categories.
[b] Percentages of trials out of 64 per age group.
[c] Percentages of trials out of 128 per age group.

Table 4
Character's Mind, Attributes, and Preferences Explanations by Age Group and Entity Type[a]

	Age group			
Entity type	3-year-olds	5-year-olds	7-year-olds	Adults
Character's mind				
Real[b]	4	35	29	56
Imaginary[b]	8	50	50	70
Total[c]	6	43	39	63
Character's preferences				
Real[b]	24	7	5	4
Imaginary[b]	21	3	1	2
Total[c]	22	5	3	3
Character's attributes				
Real[b]	22	33	31	27
Imaginary[b]	23	21	24	35
Total[c]	23	27	27	30

[a] All numbers are percentages of trials. Note that percentages for each group do not add up to 100% because some explanations did not fit any of the categories.
[b] Percentages of trials out of 64 per age group.
[c] Percentages of trials out of 128 per age group.

4.61, $p < .01$, $\eta_p^2 = .19$, with 3-year-olds providing significantly more preference explanations compared to 5- and 7-year-olds and adults, respectively. Findings for attribute explanations were null.

Explaining own predictions versus explaining different emotions. Recall that during the last four story trials, after participants offered their own judgments about story characters' likely emotions, the experimenter explicitly told and showed them that the two characters had fear reactions that differed in intensity (e.g., "Look! The baby is less afraid than Sue"; Figure 3). We compared participants' emotion explanations for these prompted trials to their responses when they explained emotions that they had predicted for story characters. Three separate repeated measures ANOVAs were performed: the first on the number of trials stimulus-oriented explanations were offered exclusively, the second on the number of trials both stimulus- and person-oriented explanations were offered together, and the third on the specific subcategory of person mind explanations. For all analyses, character (3: child, grownup, and baby) and question type (2: explain-own prediction vs. explain-different emotions) were the within-subjects variables and age group (4: 3-, 5-, and 7-year-olds and adults) was the between-groups variable.

All three analyses yielded main effects for question type, $F(1, 51) = 26.19$, $\eta_p^2 = .32$; $F(1, 55) = 17.01$, $\eta_p^2 = .26$; and $F(1, 55) = 8.4$, $\eta_p^2 = .13$, respectively. Participants of all age groups gave exclusive stimulus-oriented explanations significantly less often (16% vs. 30% of trials) and combined stimulus- and person-oriented explanations significantly more often (78% vs. 66% of trials) for trials where they were directly asked to explain why two characters had different emotions. Moreover, participants of all ages explained emotions in relation to characters' cognitive mental states significantly more often in explain-different emotion versus explain-own prediction trials (48% vs. 38% trials). Main effects for age group also emerged in these analyses (as discussed earlier).

Other Findings

At the end of the interview, we asked participants to self-report their own fears of the real and imaginary stimuli used in the stories. Comparing children's own fear reactions to those they attributed to the child characters in the stories, we found that 3-year-olds predicted more intense fears for the child protagonists than for themselves ($M = 1.00$, $SE = 0.08$ vs. $M = 0.15$, $SE = 0.21$), $F(1, 15) = 15.81$, $\eta_p^2 = .51$. In contrast, 5-year-olds and adults attributed the same level of fear to themselves ($M = 1.15$, $SE = 0.22$ and $M = 0.96$, $SE = 0.14$, respectively) as to the same-aged protagonists (i.e., child or grownup; $M = 1.33$, $SE = 0.23$ and $M = 0.89$, $SE = 0.19$, respectively); $Fs < 3.01$. Seven-year-olds provided matched self-other predictions in scenarios featuring real creatures; however, they predicted significantly less intense fears of imaginary creatures for themselves ($M = 0.29$, $SE = 0.19$) than for the child protagonists ($M = 1.09$, $SE = 0.23$), $F(3, 60) = 3.37$, $\eta_p^2 = .14$.

Finally, we questioned participants about their belief regarding the reality–fantasy distinction. As shown in Table 5, almost two thirds of the 3-year-olds insisted that at least one of the four imaginary creatures exist, with ghosts, witches, and monsters (in descending order) being the most commonly endorsed creatures ($M = 2.2$ out of the four imaginary creatures were endorsed as real by 3-year-olds; range = 0–4). On the other hand, only one 5-year-old claimed that these creatures exist (a ghost). Judgments that one or more of the imaginary creatures is or could be real (*could be* = the participant answered "maybe" to the "yes–no" question about whether the creature existed) increased at age 7 ($M = 44\%$) and continued to rise to adulthood ($M = 63\%$). Seven-year-olds and adults most frequently endorsed the real or potential existence of ghosts and then witches ($M = 1.29$ imaginary creatures endorsed by 7-year-olds, range = 0–2; $M = 1.5$ for adults, range = 0–2). Chi-square analyses confirmed significant age differences in beliefs that imaginary creatures "maybe" or "do" exist, $\chi^2(3) = 13.71$, $p < .01$, with a significant decline between 3 and 5 years and a significant increase between 5 and 7 years. Note that these data do not include interpretations of witches as humans who merely identify themselves as witches, dragons

Table 5

Percentage of Participants Endorsing the Possible Existence of Imaginary Creatures

	Age group			
Does it exist?	3-year-olds	5-year-olds	7-year-olds	Adults
"Yes"[a]	63	6	19	31
"Maybe"[a]	0	0	25	38
Either "yes" or "maybe"[a]	63	6	44	63
"No" to all[b]	37	94	56	37

[a] All numbers are percentages of participants who endorsed the existence or possible existence of at least one of the four imaginary creatures (ghosts, witches, monsters, and/or dragons).

[b] All numbers are percentages of participants who said that all four imaginary creatures do not exist.

as komodo dragons, monsters as deformed real animals, or ghosts as people dressed up.

Discussion

Results of the current study reveal that between the ages of 3 and 7 years, children develop increasing awareness that infants, children, and adults can have different emotional reactions to the same potentially fear-inducing situations. Developmental changes in fear predictions were accompanied by age-related increases in attention to character's *minds* (i.e., beliefs and knowledge) as the cause of person's fear or lack of fear, as opposed to the stimulus itself. Such mental explanations were particularly prevalent when explaining characters' fears of imaginary versus real creatures. These data advance current knowledge about the development of children's understanding of mind and emotion during early childhood.

Predictions of Fear

Children and adults predicted whether infant, child, and adult protagonists in the same situation would feel afraid versus not afraid and then made judgments about the intensity of characters' fears (if afraid). Results showed that young children as well as adults predicted that one character would feel afraid and the other character would experience no fear only 50% of the time (chance level). Significant age differences emerged, however, in participants' predictions of fear intensity. Between the ages of 3 and 7 years, children increasingly predicted different levels of fear for different-aged characters, with 5- and 7-year-olds and adults predicting different fear intensities for different characters significantly above chance for all story types (3-year-olds were above chance just for imaginary creature trials). Thus, 5- and 7-year-olds as well as adults, most frequently judged that people of different ages in the same potentially fear-inducing situation would experience emotions that differed not in absolute type (one afraid and the other not afraid) but rather in the relative *degree* of fear.

Significant age differences also appeared in the specific fear intensity predictions children and adults provided: both in relation to the character experiencing the emotion and in relation to whether the entity was real versus imaginary. First, participants of all age groups, even 3-year-olds, judged that child-aged protagonists would be significantly more afraid than grownups in situations involving a potential threat from a real creature. Importantly, however, 3- and 5-year-olds thought babies would be significantly more afraid than children, 7-year-olds thought babies and children would experience the same level of fear, and adults predicted that both babies and adults would experience less fear than child-aged protagonists. These data confirm and extend previous studies showing that 3- to 8-year-olds typically overestimate, or are simply confused about, what babies know and can do (Dowker et al.,1994; Miller et al., 2003; Montgomery, 1993; Taylor et al., 1991). In many respects, young children's confusion about the emotions of infants is not surprising. That is, children have much more limited exposure to babies in comparison to adults and same-aged peers, and they may hear contrasting opinions about infant abilities from people around them. Indeed, even adults from different cultural and social classes have divergent ideas about infants' cognitive and emotional capacities (Reznick, 1999; Reznick & Schwartz, 2001).

Finally, significant developmental changes were also apparent in participants' judgments about whether the type of potentially fearsome entity would cause different fear reactions. Whereas 3- and 5-year-olds did not differentiate fear predictions by the reality status of the entity, 7-year-olds and adults predicted that people's fears of imaginary creatures would be more intense. Still, aside from the 7-year-old group, all age groups' judgments about the relative intensity rankings of different age groups' fears were equivalent across scenarios featuring real or imaginary creatures. That is, 3- and 5-year-olds expected babies to be most afraid and adults to be least afraid in all situations, and adults expected child-aged protagonists to be more afraid than babies or adults of both real and imaginary creatures. Curiously, although 7-year-olds predicted that infants and adults would show *less* fear than child protagonists in situations featuring real creatures (same as adults), they predicted that infants and adults would experience *more* intense fear than child protagonists in scenarios featuring potential threats from imaginary creatures (opposite to the adult pattern). We will reserve interpreting this peculiar finding for the 7-year-olds until discussing results for children's predictions about their own fears of the real and the imaginary.

Explaining the Causes of Fear

Participants' explanations for emotions provide a deeper look into the development of children's knowledge about the causes of fear. Between the ages of 3 and 7 years, children offered significantly fewer explanations that pertained just to characteristics or capabilities of the stimulus (e.g., "The bee can sting him" and "The dragon can blow fire at them") and significantly more explanations that considered both the person (characteristics, preferences, and capabilities) and the stimulus simultaneously (e.g., "He *knows* that if he stands still the bee won't sting him"). Even within these combined person + stimulus explanations, there were important age differences. Whereas 3- and 5-year-olds typically focused on people's physical attributes (e.g., "Because the baby is tiny") or preferences (e.g., "Because she doesn't like snakes"), 7-year-olds and adults more often focused on character's cognitive mental states (e.g., "The baby is too young to *realize* what this thing is"). Thus, with increasing age, children evidenced greater knowledge that it is not the stimulus per se that induces fear but that fear reactions are mediated by the person's mind, especially his or her thoughts, beliefs, and knowledge. This finding challenges previous conclusions that preschoolers are mere "situationists" in that they base emotion predictions on objective features of the situation alone (e.g., Gnepp, 1989; Gnepp, Klayman, & Trabasso, 1982) and confirm more recent data suggesting that even young children are cognizant of mental causes of emotions (e.g., Lagattuta, 2007; Lagattuta & Wellman, 2001; Lagattuta et al., 1997; Pons, Harris, & de Rosnay, 2004; Rieffe et al., 2005).

Interestingly, all age groups, including 3-year-olds, referred significantly more often to the characters' cognitive mental states when explaining emotions in scenarios involving potential threats from imaginary versus real creatures. This earlier precocity in reasoning about mind – fear connections in the context of imaginary threats fits well with previous research showing that preschoolers' heightened fears often result from thinking about or imagining scary things (Muris et al., 1997, 2000). Indeed, during this age period, children spend a considerable amount of time pretending, play that often involves supernatural beings (Woolley, 1997). Children's direct experience of becoming afraid after thinking about or imagining supernatural creatures may contribute to their earlier recognition of the influence of the mind on fears of dragons, witches, ghosts, and monsters. Children's enhanced knowledge may also stem from direct adult instruction. That is, in their attempts to calm their children's fears, parents likely explain that even though people can imagine things like monsters in their minds, these creatures are not real. These assumptions regarding connections between parent – child discourse and children's knowledge about mental causes of fear need to be explored in future research before stronger conclusions can be made.

Still, it is important to consider that it may not be the inclusion of supernatural beings per se that provoked children and adults to provide significantly more mental explanations for emotions in the imaginary versus real creature scenarios. Even though both real and imaginary stimuli were drawn ambiguously (the snake could be a worm and the ghost could be a light) and the phrase *looks like* (a snake, a ghost, etc.) was included across all trial types, scenarios involving imaginary creatures were, by their very nature, more *uncertain* situations. That is, if something looks like a snake, it could really be a snake, but if it looks like a dragon, but dragons do not exist, then the identity of the creature (and one's emotional reaction to it) is more dependent upon a person's knowledge, beliefs, and expectations. Recall that Rieffe et al. (2005) found that 4-year-olds frequently provided mental explanations for a character's fear: The scenario they used involved the uncertain situation of a person entering a dark room occupied by an unknown person. Thus, future research is needed to examine whether supernatural status or just general uncertainty about the identity of a potentially dangerous entity elicits more frequent mental explanations for fear from children and adults.

Finally, results from explanation analyses showed that asking children to explain differences in people's reactions ("Look! The baby is less afraid than the child!") enabled young children to demonstrate stronger knowledge that people's minds can mediate between situations and emotions. These explain-different emotion trials provoked significantly more cognitive mental state explanations for emotions (e.g., "The grown-up knows that ghosts are not real") than the *predict-and-explain* trials. These data converge with recent findings that asking young children to explain divergent or unexpected behaviors or emotions can be an effective tool for uncovering more sophisticated reasoning about people's minds in young children (Bartsch & Wellman, 1995; Lagattuta 2005; Robinson & Mitchell, 1995; Tardif, Wellman, & Cheung, 2004). Still, it is important to acknowledge that even with the inclusion of an explicit prompt to explain divergent emotions, the majority of 3-year-olds' emotion explanations were stimulus- rather than person oriented. Thus, 3-year-olds' understanding about mind – emotion connections is clearly more fragile and less coherent than that of older children and adults.

Self Versus Others and Ontological Reasoning

When asked to rate the intensity of their own fears of the creatures included in the story scenarios (at the end of the interview), 5-year-olds and adults matched their personal fear ratings with those they predicted for the same-aged characters in the scenarios. In contrast, 3-year-olds predicted significantly higher fears for child-aged protagonists than for themselves, and 7-year-old predicted significantly more intense fears of imaginary (but not real) creatures for child-aged protagonists than for themselves. If children make sense of others' emotions by considering how they would feel in that situation (Harris, 1989), then why did 3- and 7-year-olds make different emotion reports for themselves than for same-aged story characters?

Previous research has shown that preschool-aged children often evaluate themselves in a highly positive and ideal manner (Harter, 1999) and strive to increase their ability to act autonomously or be a "big kid" (see Lagattuta & Thompson, 2007, for a review). Because fear can be regarded as a sign of weakness or babyishness, 3-year-olds may have exaggerated their own personal courageousness while at the same time acknowledged that children of their same age do experience fear. Alternatively, when asked about their own fears, 3-year-olds might simply not be aware that they indeed are afraid of these creatures.

Divergence in self versus other emotions predictions for 7-year-olds may result from different developmental processes. Starting around the ages of 7 – 9 years, children begin to engage in more hypothetical or counterfactual thinking as well as show stronger knowledge that people can have different interpretations about what is real (Siegler & Alibali, 2005). These cognitive advances may lead children to begin questioning their assumptions that supernatural creatures do not exist. Indeed, as shown in our data, there was a significant increase between 5 years (6% participants) and 7 years (44% participants) in judging that one or more of the imaginary creatures (especially ghosts) are or could be real. This remained high even for adult participants (63% participants; see also Rice, 2003, for evidence that many adults endorse the existence of supernatural creatures). These data fit with recent findings from Harris, Pasquini, Duke, Asscher, and Pons (2006) that between 4 and 8 years of age, children become increasingly aware that people hold mixed opinions about whether "equivocal beings" such as ghosts, monsters, and angels exist in real life.

Because 7-year-olds are transitioning into a period of questioning the possibility of the supernatural, this may have caused their reasoning on the imaginary creature trials to be nuanced and inconsistent compared to the other age groups' responses. Seven-year-olds predicted higher levels of fear of imaginary creatures for child-aged protagonists than for themselves (but made corresponding self-other predictions for real creatures), and they predicted that adults and babies would be more afraid than children when encountering something that looks like a ghost, witch, dragon, or monster. Still, even these seemingly confused predictions were often explained on very reasonable grounds: "He (the grownup) knows not to believe in ghosts, but he sees one, so he is really, *really* scared!" Future research needs to more systematically explore how older children and young adults learn to reconcile opposing viewpoints about the ontological status of supernatural creatures from the mixed testimony and evidence they receive from peers, adults, movies, and stories (Harris & Koenig, 2006). For example, parents may claim, "Monsters are not real," yet the older child may read stories or watch movies with very realistic monsters. Indeed, further exploration into a potential developmental transition between 5 and 7 years from *not* believing to believing in the possible existence of the supernatural is warranted.

Future Research and Conclusions

Results of the current study show that between the ages of 3 and 7 years, children increasingly understand that people of different ages can experience different emotions in the same situation. Yet, this still does not clarify whether children believe that dissimilar fear reactions of the two individuals stem from group-specific or person-specific differences (Gnepp, 1989). That is, each scenario contrasted two individuals from separate groups (i.e., grownup vs. child or child vs. infant) but did not contrast two individuals from the same group (i.e., two children, two grownups, or two infants). This conflated the individual with the group, and, as a result, participants' judgments regarding differences between the two story characters cannot conclusively be attributed to either group-specific or person-specific factors. Thus, more needs to be explored about the extent to which children think certain groups may be more or less homogenous in the way they interpret and emotionally react to situations.

Second, previous research has shown that it is not until the age of 7 or 8 that children begin to suggest mental strategies for coping with negative emotions, including fear (e.g., Altshuler & Ruble, 1989; Band & Weisz, 1988). Because the current data reveal that children view fears of imaginary creatures as more causally connected to people's cognitive mental states than fears of real creatures, it is possible that children's initial insights into cognitive strategies for alleviating fear may be in the context of coping with fears of imaginary beings like ghosts and witches. Indeed, it would be interesting to expand on the current study to examine children's suggestions of ways to deal with fears to assess whether young children have stronger knowledge about how to use the mind to cope with fears of the imaginary (e.g., pretending an imagined ghost to be friendly instead of mean) versus real threat situations.

In conclusion, between the ages of 3 and 7 years, children develop more sophisticated knowledge that people of different age groups can experience different intensities of fear in the same situation. Indeed, with increasing age, children demonstrated a stronger appreciation of the mind as mediating between situations and emotions—people's fears depend on their thoughts, beliefs, and knowledge, not just the situation itself—with this knowledge about mental causes of fear being most robust in the context of fears of the imaginary. Future research needs to systematically examine how children's beliefs, questions, and the testimony they hear about what should and should not be feared, as well as what is real and not real, influence the frequency and intensity of their fear in their everyday lives.

References

Altshuler, J. L., & Ruble, D. N. (1989). Developmental changes in children's awareness of strategies for coping with uncontrollable stress. *Child Development*, *60*, 1337–1349.

Band, E. B., & Weisz, J. R. (1988). How to feel better when it feels bad: Children's perspectives on coping with everyday stress. *Developmental Psychology*, *24*, 247–253.

Bartsch, K., & Wellman, H. M. (1995). *Children talk about the mind*. Oxford, England: Oxford University Press.

Denham, S. A. (1986). Social cognition, prosocial behavior, and emotion in preschoolers: Contextual validation. *Child Development*, *57*, 194–201.

Dowker, A., Hart, A., Heal, A., Phillips, C., & Wilson, A. (1994, December). *"Developmental psychologists" in the reception class: Young children's ideas about age differences in abilities*. Poster session presented at the British Psychological Society Developmental Section Conference, Portsmouth, England.

Ekman, P., Sorenson, E. R., & Friesen, W. V. (1969). Pancultural elements in facial displays of emotion. *Science*, *164*, 86–88.

Gnepp, J. (1983). Children's social sensitivity: Inferring emotions from conflicting cues. *Developmental Psychology*, *19*, 805–814.

Gnepp, J. (1989). Personalized inferences of emotions and appraisals: Component processes and correlates. *Developmental Psychology*, *25*, 277–288.

Gnepp, J., & Gould, M. E. (1985). The development of personalized inferences: Understanding other people's emotional reactions in light of their prior experiences. *Child Development*, *56*, 1455–1464.

Gnepp, J., Klayman, J., & Trabasso, T. (1982). A hierarchy of information sources for inferring emotional reactions. *Journal of Experimental Child Psychology*, *33*, 111–123.

Gullone, E. (2000). The development of normal fear: A century of research. *Clinical Psychology Review*, *20*, 429–451.

Harris, P. L. (1989). *Children and emotion: The development of psychological understanding*. Cambridge, MA: Basil Blackwell.

Harris, P. L., & Koenig, M. A. (2006). Trust in testimony: How children learn about science and religion. *Child Development*, *77*, 505–524.

Harris, P. L., Pasquini, E. S., Duke, S., Asscher, J. J., & Pons, F. (2006). Germs and angels: The role of testimony in young children's ontology. *Developmental Science*, *9*, 76–96.

Harter, S. (1999). *The construction of the self: A developmental perspective*. New York: Guilford.

Lagattuta, K. H. (2007). Thinking about the future because of the past: Young children's knowledge about the causes of worry and preventative decisions. *Child Development*, *78*, 1492–1509.

Lagattuta, K. H. (2005). When you shouldn't do what you want to do: Young children's understanding of desires, rules, and emotions. *Child Development*, *76*, 713–733.

Lagattuta, K. H., & Thompson, R. A. (2007). The development of self-conscious emotions: Cognitive processes and social influences. In J. L. Tracy, R. W. Robins, & J. P. Tangney (Eds.), *The self-conscious emotions: Theory and research* (pp. 91–113). New York: Guilford.

Lagattuta, K. H., & Wellman, H. M. (2001). Thinking about the past: Early knowledge about links between prior experience, thinking, and emotion. *Child Development*, *72*, 82–102.

Lagattuta, K. H., Wellman, H. M., & Flavell, J. H. (1997). Preschoolers' understanding of the link between thinking and feeling: Cognitive cuing and emotional change. *Child Development*, *68*, 1081–1104.

Miller, S. A. (2000). Children's understanding of preexisting differences in knowledge and belief. *Developmental Review*, *20*, 227–282.

Miller, S. A., Hardin, C. A., & Montgomery, D. E. (2003). Young children's understanding of the conditions for knowledge acquisition. *Journal of Cognition and Development*, *4*, 325–356.

Montgomery, D. E. (1993). Young children's understanding of interpretive diversity between different-age listeners. *Developmental Psychology*, *29*, 337–345.

Muris, P., & Merckelbach, H. (2000). How serious are common childhood fears? The parent's point of view. *Behaviour Research and Therapy*, *38*, 813–818.

Muris, P., Merckelbach, H., & Collaris, R. (1997). Common childhood fears and their origins. *Behaviour Research and Therapy*, *35*, 929–937.

Muris, P., Merckelbach, H., Gadet, B., & Moulaert, V. (2000). Fears, worries, and scary dreams in 4- to 12-year-old children: Their content, developmental pattern, and origins. *Journal of Clinical Child Psychology*, *29*, 43–52.

Oatley, K., & Jenkins, J. M. (1996). *Understanding emotions.* Malden, MA: Blackwell.

Palermo, D. S. (1974). Still more about the comprehension of 'less.' *Developmental Psychology, 10,* 827–829.

Phelps, K. E. (1999). *Young children's understanding of fears.* Unpublished dissertation, University of Texas, Austin.

Pons, F., Harris, P. L., & de Rosnay, M. (2004). Emotion comprehension between 3 and 11 years: Developmental periods and hierarchical organization. *European Journal of Developmental Psychology, 1,* 127–152.

Rieffe, C., Meerum Terwogt, M., & Cowan, R. (2005). Children's understanding of mental states as causes of emotions. *Infant and Child Development, 14,* 259–272.

Reznick, J. S. (1999). Influences on maternal attribution of infant intentionality. In P. D. Zelazo, J. W. Astington, & D. R. Olson (Eds.), *Developing theories of intention: Social understanding and self control* (pp. 243–267). Mahwah, NJ: Erlbaum.

Reznick, J. S., & Schwartz, B. B. (2001). When is an assessment an intervention? Parent perception of infant intentionality and language. *Journal of the American Academy of Child & Adolescent Psychiatry, 40,* 11–17.

Rice, T. W. (2003). Believe it or not: Religious and other paranormal beliefs in the United States. *Journal for Scientific Study of Religion, 42,* 95–106.

Robinson, E. J., & Mitchell, P. (1995). Masking of children's early understanding of the representational mind: Backwards explanation versus prediction. *Child Development, 66,* 1022–1039.

Rosengren, K. S., Johnson, C. N., & Harris, P. M. (2000). Imagining the impossible: Magical, scientific, and religious thinking in children. *Journal of the American Academy of Child & Adolescent Psychiatry, 40,* 1484–1486.

Sarafino, E. P. (1986). *The fears of childhood: A guide to recognizing and reducing fearful states in children.* New York: Human Sciences Press.

Siegler, R. S., & Alibali, M. W. (2005). *Children's thinking* (4th ed.). Upper Saddle River, NJ: Prentice Hall.

Tardif, T., Wellman, H. M., & Cheung, K. M. (2004). False belief understanding in Cantonese-speaking children. *Journal of Child Language, 31,* 779–800.

Taylor, M., Cartwright, B. S., & Bowden, T. (1991). Perspective taking and theory of mind: Do children predict interpretive diversity as a function of differences in observers' knowledge? *Child Development, 62,* 1334–1351.

Thompson, R. A., & Lagattuta, K. H. (2006). Feeling and understanding: Early emotional development. In K. McCartney & D. Phillips (Eds.), *Blackwell handbook of early childhood development* (pp. 317–337). Malden, MA: Blackwell.

Wannemacher, J. T., & Ryan, M. L. (1978). "Less" is not "more": A study of children's comprehension of "less" in various task contexts. *Child Development, 49,* 660–668.

Wenar, C. (1990). Childhood fears and phobias. In M. Lewis & S. M. Miller (Eds.), *Handbook of developmental psychopathology* (pp. 281–290). New York: Plenum Press.

Woolley, J. D. (1997). Thinking about fantasy: Are children fundamentally different thinkers and believers from adults? *Child Development, 68,* 991–1011.

Child Development, July/August 2008, Volume 79, Number 4, Pages 836–851

Gender and Patterns of Concerned Responsiveness in Representations of the Mother–Daughter and Mother–Son Relationship

Ruth Butler and Rachel Shalit-Naggar
Hebrew University of Jerusalem

Given that girls show more interpersonal concern than boys, it was predicted that more mother–daughter than mother–son dyads would develop a relationship of mutual concerned responsiveness (CR). Two hundred and twenty-six Israeli children (7–8 years old) and 91 mother–child pairs provided narratives of mother–child interactions. At high levels of socioeconomic status (SES), descriptions of child but not maternal concern differed by gender; therefore, more mother–daughter narratives described mutual CR and more mother–son narratives described a nonreciprocal pattern of maternal CR. In a low-SES sample, most mother–daughter narratives described mutual CR, but many mothers and sons described little concern by either partner. Results provided clear evidence of gender differences in mother–child reciprocity and confirmed the importance of examining gender influences in different social groups.

There is still active debate whether and how mother–daughter and mother–son interactions differ in childhood (Leaper, 2002; Lytton & Romney, 1991; Siegal, 1987). In this article, we examined the novel proposal that one significant way in which they differ is in the degree to which both partners show active concern for promoting the welfare of the other and the relationship. Prior studies have not shown gender differences in mother–child mutual responsiveness (Maccoby, 1998). To date, however, researchers have focused mainly on the role of maternal sensitive concern in creating a reciprocal relational system (Kochanska, 1997). In contrast, we examined narrative representations of concern on the part of both mother and child. The first main prediction was that daughters would be described as displaying more active concern than sons. Because mothers tend to display similar levels of concern toward sons and daughters (Leaper, 2002), we thus predicted that more mother–daughter than mother–son narratives would describe a pattern of mutual concerned responsiveness (CR), and more mother–son narratives would describe interactions characterized by concern on the part of the mother alone.

Gender, Concern, and Patterns of Responsiveness in the Mother–Child Relationship

Girls display higher levels of empathic concern, prosocial behavior, and interpersonal responsibility than do boys, across a variety of measures and contexts (see Eisenberg & Fabes, 1998; Maccoby, 1998, for reviews). These gender differences also have implications for boys' and girls' relationships. In their analysis of same-sex peer groups, Maltz and Borker (1982) distinguished between the collaborative, affiliative speech of girls who speak to "create and maintain relationships of closeness and equality" (p. 207) and the competitive, adversarial speech of boys who speak to assert themselves and to maintain an audience. Studies of same-sex friendships in later life have yielded similar findings, whereby women were more oriented than men to understanding and caring for others and to maintaining relational harmony and mutuality (Burke & Fuqua, 1987; Wright, 1982). Moreover, when the sexes meet, women tend to engage more than their male partners in support work that serves to encourage and maintain communication by active listening and sensitive responding (Fishman, 1983; Tannen, 1993). More generally, Bartky (1990) has described heterosexual relationships as characterized by an unequal division of "emotional labor," whereby women invest more effort than do men to promote their partner's emotional welfare.

In a similar vein, some theorists have proposed that the mother–daughter relationship is more likely to

---

Rachel Shalit-Naggar is now a clinical child psychologist in private practice.

Study 1 was based in part on the master's thesis of the second author. We would like to thank Lydia Ginsburg and Naama Navin for helping to coordinate data collection for Study 2, the many undergraduate students who contributed to data collection and coding, and, above all, the children and mothers who participated in this research.

Correspondence concerning this article should be addressed to Ruth Butler, School of Education, Hebrew University of Jerusalem, Jerusalem 91905, Israel. Electronic mail may be sent to msbutler@mscc.huji.ac.il.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7904-0003

develop along a trajectory of reciprocal responsiveness than the mother–son relationship because mothers are predisposed to create more intimate relations with daughters (Chodorow, 1978; Gilligan, 1996). Analyses of mother–child mutual responsiveness have also emphasized the contribution of the mother and, in particular, the role of maternal sensitivity in inducting children into a system of reciprocity (Kochanska, 1997; Maccoby & Martin, 1983; Maccoby & Parpal, 1985). Originally developed in the context of attachment theory, maternal sensitivity refers to the degree to which the mother is attuned to the child's needs and responds to them in ways intended to promote the child's welfare and to maintain a warm mother–child relationship (Bell & Ainsworth, 1972).

One possibility is that mothers may be more likely to induct daughters into a mutually responsive relationship because they tend to respond more sensitively to daughters than to sons. Most studies have not shown gender differences in the level of observed or reported maternal sensitivity, however (see Leaper, 2002; Lytton & Romney, 1991, for reviews), and the few exceptions have yielded mixed findings as to whether mothers responded more sensitively toward daughters (e.g., Barnett, Kidwell, & Leung, 1998) or sons (e.g., Dodge, Petit, & Bares, 1994). Moreover, studies of children's narrative representations have not shown differences in the degree to which girls and boys themselves described the mother as warm and responsive (Oppenheim, Emde, & Warren, 1997; Solomonica-Levi, Yirmiya, Erel, Samet, & Oppenheim, 2001). Most significantly, in the present context, Maccoby (1998) concluded that studies have not shown gender differences in mother–child mutual responsiveness.

The main objective of the present studies was to examine the proposal that differences in mutual responsiveness in mother–son versus mother–daughter dyads will emerge more clearly than hitherto if one considers possible differences in the contributions of girls and boys to the interactive arena. Specifically, we conceptualized a dimension of child active concern (CAC), analogous in some senses to that of maternal sensitivity, to capture individual differences in the degree to which children respond to the mother's needs and feelings in ways that reflect concern to promote her welfare and to maintain relational harmony. Extrapolating from evidence of early gender differences in prosocial behavior, we predicted that girls will exhibit more such active concern than will boys in the mother–child as in other relationships.

At first sight, this conceptualization may seem inappropriate, given the asymmetrical nature of the mother–child relationship, especially during childhood. Indeed, to date, researchers have defined child responsiveness to the mother mainly in terms of children's positive cooperation and receptive compliance with maternal signals, initiatives, and influence, rather than their proactive concern for the mother and the relationship (Biringen & Robinson, 1991; Kochanska, 1997; Maccoby & Martin, 1983). Studies have shown, however, that already during the first 3 years, some children responded to maternal distress or anger by spontaneously offering comfort or help or trying to repair ruptures in the relationship (Kochanska, Forman, & Coy, 1999; Radke-Yarrow, Zahn-Waxler, Richardson, Susman, & Martinez, 1994; Zahn-Waxler, Radke-Yarrow, Wagner, & Chapman, 1992). Moreover, such responses were more common among girls than boys.

These findings support our proposal that girls are more likely than boys to display active concern within the context of the mother–child relationship. This proposal has clear implications for mutual responsiveness in mother–daughter and mother–son dyads. Specifically, if both mothers and children differ in the degree to which they are oriented to promote the welfare of the other and the relationship, one can theoretically posit four distinct patterns: mutual CR, whereby both mother and child display active concern; maternal CR, whereby the mother but not the child displays active concern; child CR; and no CR by either partner. Moreover, if child gender influences CAC more than it does maternal active concern (MAC), one can predict that more mother–daughter than mother–son dyads will display a pattern of mutual CR, typical also of close female friendships. In contrast, more mother–son dyads should display a pattern of maternal CR, reminiscent of the unequal division of emotional labor that has emerged in studies of other cross-sex interactions.

Although some other theorists have also proposed that mothers are more likely to develop mutually binding relationships with daughters than sons (Chodorow, 1978; Leaper, Anderson, & Sanders, 1998; Maccoby, 1990), Maccoby (1998) concluded that studies have not shown greater mutuality in mother–daughter than in mother–son dyads in childhood. The present analysis suggests that this may have been because researchers did not examine child and thus reciprocal concern. In keeping with prevailing conceptualizations of child responsiveness, Kochanska (1997) defined a "mutually responsive orientation" when dyads show high levels of maternal sensitivity, child committed compliance, and positive mother–child interactions; others have defined reciprocity mainly in terms of shared positive affect (Ainsworth, Blehar, Waters, & Wall, 1978; Dix, 1991). Studies have not shown differences between mother–daughter and mother–son dyads in these kinds of mutual responsiveness (Kochanska, 1997, 2002; Maccoby, 1998). In contrast, we predicted that examining the degree to which both partners are oriented to promote the welfare of the other and the relationship will confirm that more mother–daughter than mother–son dyads will experience a relationship of "mutual CR."

Examining Active Concern and Patterns of CR in the Mother–Child Relationship

The use of projective methods to provide a window on people's emotional and interpersonal worlds has a long history in psychology. For example, in the context of research on the mother–child relationship, studies have shown that children's narratives could be reliably scored for descriptions of maternal sensitive concern that were coherently associated with other measures of child and maternal behavior and adjustment (Oppenheim et al., 1997; Solomonica-Levi et al., 2001). In the present studies, children told stories about pictures in which a mother and her child were depicted in contexts conducive to eliciting concern for the welfare of the other and the relationship from both partners. Thus, in some pictures, the mother was angry, sad, or sick, and in others, the child was sad or angry. Narratives were scored in two ways. First, aggregate measures of maternal and CAC were derived by counting the number of stories in which the mother or the child, respectively, was described as acting to promote the welfare of the other and the relationship. Second, entire protocols were assigned to one or another category of CR based on the degree to which the global impression over all stories and situations was of mutual, maternal, child, or no CR. Children also told a story about an emotionally neutral picture, in which mother and child were depicted sitting on a bench; this story was scored for positive emotional climate. Inclusion of this measure was designed to provide some point of comparison with prior studies of mutual responsiveness. Thus, we anticipated that the well-established association between observational measures of maternal sensitivity and positive mother–child interactions (see Kochanska, 2002; Maccoby, 1998, for reviews) would be replicated also when we examined the relation between representations of MAC in children's stories about negative emotions and representations of the emotional quality of mother–child interactions in their "bench" story.

We report two studies. Participants in Study 1 were children living in a middle-class neighborhood. We focused on 7- to 8-year-olds on the assumption that by this age, children can generate stories without excessive prompting. The first objective was to verify that stories could be reliably scored for descriptions of maternal and CAC and patterns of CR. The second and major objective was to examine the predictions that (a) descriptions of CAC would be more frequent in girls' than in boys' stories, but girls and boys would describe similar levels of MAC and (b) more girls' than boys' protocols would describe a pattern of mutual CR, and more boys' protocols would describe a nonreciprocal pattern of maternal CR. In Study 2, we reexamined these predictions for a larger and more socially diverse sample that included both children and some of their mothers. The main objective was to provide additional validation for our framework by testing the predictions that gender would be related to descriptions of child and mutual concern responsiveness among both children and mothers, at both higher and lower levels of maternal education. Another objective was to examine the further prediction that representations of children and their mothers would be correlated within mother–child dyads.

Study 1

Method

Participants

Participants were 63 children (31 boys, 32 girls, mean age = 97 months, *SD* = 3.62) in two Grade 2 classes in a school serving a middle-class neighborhood in West Jerusalem. Israel has separate school systems for Hebrew-speaking (mainly Jewish) and Arabic-speaking populations; the Hebrew language system is further divided into secular and religious systems. Participants for the present studies were drawn from Hebrew language secular schools. Parents provided informed consent, but the principal did not authorize release of demographic information from students' files.

Instrument

An artist prepared six line drawings of a mother and either a boy or a girl; facial features differed in each; drawings are presented in Figure 1. Five pictures depicted negative emotional states. In each picture, one figure was drawn full-face and one in profile. In two pictures, the mother's face expressed anger or sadness, respectively, and in another two, the child's face expressed anger or sadness. The fifth picture depicted a mother in bed with a thermometer in her mouth. The sixth, "neutral," picture depicted

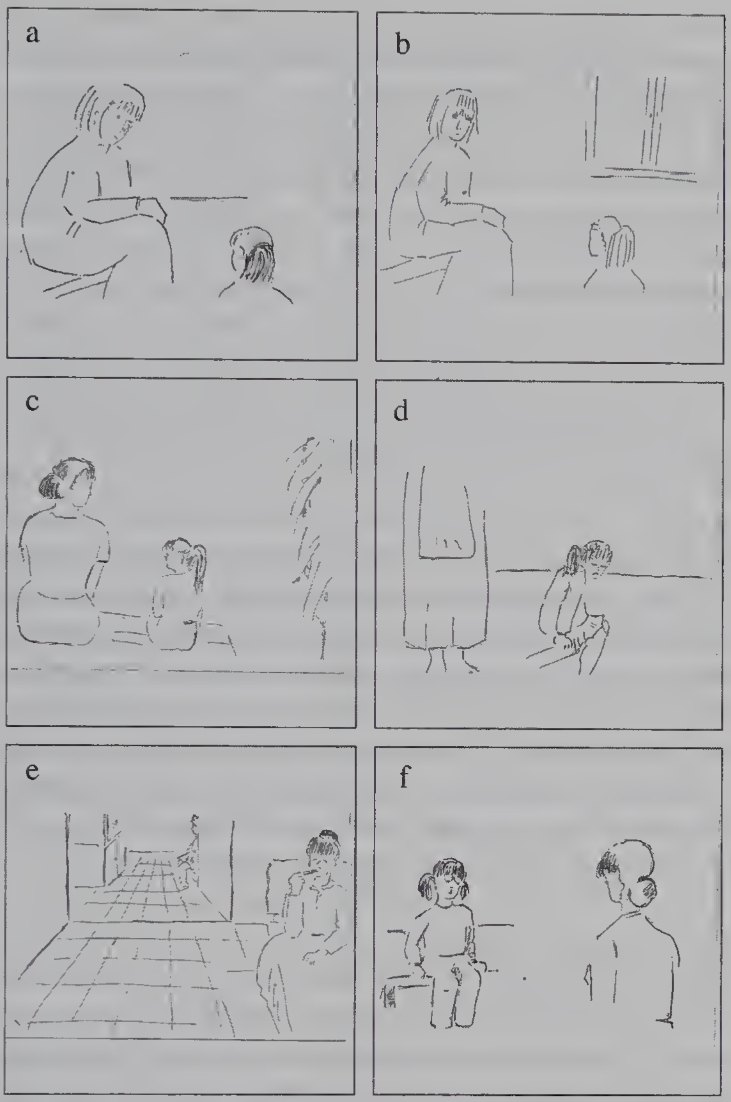


Figure 1. Drawings used to elicit mother–daughter narratives: (a) mother sad, (b) mother angry, (c) bench, (d) child sad, (e) mother sick, and (f) child angry.

a mother and a child in profile sitting on a bench. A pilot study with 15 children at ages 7–8 confirmed that all correctly identified the facial expressions and showed that administration following the sequence mother angry, mother sad, bench, child sad, mother sick, and child angry minimized carryover of themes between stories. We had intended to tape the stories, but the pilot study indicated that the tape recorder distracted the children, increased self-consciousness, and disrupted their narratives. This was not the case when administrators wrote down children's stories; they had no difficulty doing so at a pace that did not seem to affect narrative fluency.

Procedure

The teacher introduced the second author as a student interested in children's stories. The second author invited each child to a side room and presented gender-appropriate pictures in the sequence described above. For each picture, she said, "Here is a mother and her daughter (son). The child is in Grade 2. The mother (child) is angry (or sad, or the mother is sick, or they are sitting on a bench). Tell me a story about the picture: what happened before, what is happening now, what each person is feeling, saying, and doing, and what will happen next." Stories were written down verbatim; probes were used only when major elements were unclear.

Scoring

Pairs of coders who were unaware of the hypotheses scored (a) individual emotion stories, (b) entire protocols, and (c) emotional climate in the bench story. Each pair included one male and one female to control for possible gender influences in interpreting active concern as displayed by girls and boys. A third coder resolved any disagreements.

CAC and MAC. Children's stories about the five emotion pictures (mother angry, sad, sick, and child angry, sad) were each typed on a separate page; stories were scored by picture rather than by child to avoid carryover effects. Two scorers independently scored each story for the presence of CAC and MAC if the target figure acted to promote the other's welfare (e.g., comforted, helped, or tried to solve the other's difficulty) and/or to restore the relationship (e.g., apologized, made reparation). Active concern could be verbal ("The mother said she knew he was sad to miss the school trip, but she'd take him to the zoo when he got better"), nonverbal ("She hugged her mum"), or practical ("He bought his mum a vase instead of the one he broke"). Each story received a score of either 1 or 0 for both CAC and MAC. Agreement coefficients were high for both CAC (from $\kappa = .86$ for mother angry to $\kappa = .96$ for mother sad) and MAC (from $\kappa = .87$ for child angry to $\kappa = .95$ for child sad). Total CAC and MAC scores over all five stories could range from 0 to 5 and were highly correlated between judges, $r(63) = .93$ for CAC and $r(63) = .90$ for MAC.

Patterns of CR. Two scorers each read entire protocols (without the bench story) on the assumption that global readings would best capture patterns of coordination of maternal and child concern across different situations. Scorers assigned protocols to categories of CR after reading the following instructions:

> *Mutual Concerned Responsiveness* is scored when both mother and child are described as displaying an orientation of concern for the other and the relationship, rather than as guided only by their own needs, beliefs, or concerns, and as *acting* to relieve distress, to promote the other's welfare, and

to maintain or restore harmonious relations. This category applies when each partner acts responsively in different stories and/or when stories describe reciprocal concern by both. Mutual empathy (partners feel for one another, but do not act responsively) or compliance (partners just comply with one's another's directives and requests) do not qualify. *Maternal Concerned Responsiveness* is scored when the mother, but not the child is described as acting responsively as described above. Disregard other child behaviors, such as whether they cooperate with or resist maternal initiatives. *Child Concerned Responsiveness* is scored when the child, but not the mother shows active concern. *No Concerned Responsiveness* is scored when neither partner acts responsively. Instead, mother and child ignore the other's distress or ruptures in the relationship, fail to act, or act to influence or discipline the other in keeping with their own wishes, needs, or beliefs.

Scorers were instructed to consider the majority of stories if one differed markedly from the rest and to assign protocols that did not display a clear pattern to a fifth, unclear category. Agreement for assignment to the five categories was high, $\kappa = .92$.

Emotional climate. Two new scorers rated the bench story for emotional climate on a scale anchored at 5 (*very positive: mood is happy, excited*) and 1 (*very negative: mood is angry, fearful, sad*). Only two stories described conflicting emotions in mother and child; both scorers rated these 3. Ratings were highly correlated, $r(63) = .94$ and were averaged across judges.

Results

Gender, Active Concern, and Patterns of CR

As predicted, more girls' than boys' stories were scored for the presence of CAC, $t(61) = 2.10$, $p < .05$, ($M = 2.03$, $SD = 1.45$, and $M = 1.29$, $SD = 1.35$, respectively). In contrast, the effect for gender on descriptions of MAC was not significant ($M_{\text{girls}} = 1.75$, $SD = 1.27$, and $M_{\text{boys}} = 1.58$, $SD = 1.36$). We also predicted that more girls' than boys' protocols would describe a pattern of "mutual CR," and more boys' protocols would describe a pattern of "maternal CR." As expected, an initial analysis for the four categories of CR and the fifth, unclear category yielded a significant association between patterns and gender, $\chi^2(4) = 17.92$, $p < .001$ (see Table 1 for cell frequencies). Follow-up analyses compared the frequency of each pattern (coded 1) versus all other patterns (coded 0). As predicted, more girls' than boys' protocols were scored as mutual CR, $\chi^2(1) = 13.06$, $p < .001$. Indeed, 15 of the 17 such narratives were told by girls, and this was the most frequent category for girls. In contrast, more boys' than girls' protocols were scored as "maternal CR," $\chi^2(1) = 8.22$, $p < .01$, and this was the most frequent category for boys. The association with gender was not significant for the other patterns, but Table 1 shows some tendency for more boys than girls to describe a pattern of child CR.

Table 1

Percentages of Boys' and Girls' Protocols Assigned to Each Pattern of Mother–Child Responsiveness, Study 1

	Boys	Girls	All children
Mutual responsiveness	7	47	27
Maternal responsiveness	36	6	21
Child responsiveness	26	13	19
No CR	19	25	22
Unclear	13	9	11
n	31	32	63

Note. CR = concerned responsiveness

Relations Between Representations

As expected, representations of MAC were significantly correlated with positive emotional climate in the bench story for both girls, $r(32) = .51$, $p < .01$, and boys, $r(31) = .37$, $p < .05$. The level of positive emotional climate did not differ significantly by gender. In contrast, descriptions of CAC and MAC were significantly correlated among girls, $r(32) = .41$, $p < .05$, but not boys, $r(31) = -.23$; these correlations differed significantly, $z = 2.58$, $p < .01$. This was the case also for the association between CAC and positive emotional climate in the bench story, $r(32) = .42$, $p < .05$ for girls, and $r(31) = -.10$ for boys, $z = 2.08$, $p < .05$.

Discussion

Children's stories could be reliably scored for representations not only of MAC but also of CAC and patterns of CR. Most significantly, the results supported the predictions that more girls than boys would describe a relationship of reciprocal concern because of gender differences in representations of CAC. In keeping with results from studies based on observational measures of maternal sensitivity (Biringen et al., 1999; Kochanska, 1997) or a more established procedure for eliciting and scoring children's narratives (Oppenheim et al., 1997), boys and girls did not differ in the frequency with which they described the mother as acting to promote the welfare of the child and the relationship. In contrast, girls described more instances of CAC than boys. Thus, far more girls than boys described a relationship of "mutual CR" in which both partners displayed active concern, and more boys than girls described a nonreciprocal pattern of "maternal CR."

These results are quite exciting in view of Maccoby's (1998) conclusion that studies have not shown greater reciprocity in mother – daughter than in mother – son dyads in childhood. Some additional findings provided further support for our interpretation that gender is related to some, more than other, kinds of mother – child reciprocal responsiveness. In keeping with results from the studies of reciprocal positive affect reviewed by Maccoby (1998), gender did not influence descriptions of positive emotional climate in the bench story or the positive association between positive climate and representations of maternal concern. This similarity with prior findings inspires confidence in our narrative measures. In contrast, although studies have shown that maternal sensitivity was associated with committed compliance in children of both sexes (Kochanska, 1997, 2002), representations of maternal and child concern were positively correlated only in girls' narratives. Thus, there might be differences between girls and boys not only in the level of child concern but also in the degree to which children respond to maternal concern with both concern and compliance or mainly with compliance alone.

It is premature to discuss these results further because Study 1 had several limitations that we addressed in Study 2. First, the sample was small and socially homogenous. Second, claims about gender differences in the mother – child relationship necessitate incorporation also of the mother's perspective. Third, Study 1 did not provide any external validation for children's representations. Fourth, our interpretation that gender might moderate the relation between maternal and child concern more than it does the relation between maternal concern and child committed compliance must be treated with caution because we did not include a measure of committed compliance.

Study 2

In Study 2, we examined representations of mother – child interactions in a larger, socially diverse sample that included both children and some of their mothers. The first objective was to examine whether the gender differences in children's representations of child and mutual CR would be replicated at both higher and lower levels of socioeconomic status (SES), as determined by maternal education. This endeavor is important because most studies of gender influences in the mother – child relationship have examined middle-class samples (see Leaper, 2002; Leaper & Smith, 2004; Leaper et al., 1998, for reviews). Interestingly, the few studies that have shown gender differences in maternal sensitivity targeted low-SES groups, but these studies have yielded mixed findings as to whether mothers behaved more sensitively toward sons (Biringen et al., 2000; Dodge et al., 1994) or daughters (Barnett et al., 1998). Other studies of low-SES groups in both North America and Israel did not show any effect of child gender on maternal sensitivity, however (Kiang, Moreno, & Robinson, 2004; Solomonica-Levi et al., 2001). In contrast with these mixed findings, gender differences in children's prosocial behavior have emerged across a wide range of social and cultural groups (LaFreniere et al., 2002; Whiting & Edwards, 1988), and girls responded more empathically to maternal distress than boys in the only such study that targeted a low-SES sample (Kiang et al., 2004). Thus, there were grounds for anticipating that gender would affect representations of child, more than MAC at both high and low levels of maternal education, and thus that the gender differences in patterns of CR reported in Study 1 would be replicated for both social groups.

The second objective was to examine the predictions that mothers will also describe girls as displaying more active concern than sons and will be more likely to describe a relationship of mutual CR with daughters and a relationship of maternal CR with sons. Most mothers were asked to describe recent situations in which the mother and target child were present; the mother was angry, sad, or ill; the child was angry or sad; and, for the neutral cue, a situation where both mother and child were present. The aim was to elicit rich and detailed descriptions of actual mother – child interactions, but we also had some concern that this procedure might be more likely than less direct methods to bias mothers to describe their relationship with their child too positively. Thus, some mothers were asked to tell stories about the pictures given to children; some of these mothers also described actual events. This aspect of the design enabled us to compare descriptions of MAC and CAC and patterns of CR both between mothers given different procedures and within mothers who responded to both.

No studies have examined maternal reports of child or mutual responsiveness as defined here, but there is some prior evidence that mothers rated daughters as more empathic and helpful than sons (Eberly & Montemayor, 1998; Hastings, Zahn-Waxler, Robinson, Usher, & Bridges, 2000). In addition, analyses of the mother–child relationship as a system of reciprocity (Kochanska, 1997; Maccoby & Martin, 1983) imply congruence between partners' perspectives and experiences. If boys and girls differ in their descriptions of CAC and patterns of CR, one would expect child gender to be related to maternal representations in similar ways. The third objective was to examine the related prediction that children's and mothers' representations of concern will be correlated within mother–child dyads. If the results support this prediction, they will contribute to addressing another limitation of Study 1 by providing some external validation for children's representations. We also examined the stability of children's representations by retesting some of the children after about a year.

The fourth objective was to examine whether gender moderates the relation between maternal and child concern more than it does the relation between maternal concern and child committed concern. Maternal protocols were scored both for CAC and for the degree to which the mother described the child as cooperating and eagerly complying with her initiatives and directives. Following prior findings and the results of Study 1, we predicted that representations of maternal concern would be associated with descriptions of child committed compliance for both boys and girls (Kochanska, 1997, 2002) but would be associated with representations of CAC only for girls.

Method

Participants

The child sample included 83 boys and 80 girls (mean age at testing = 95 months, *SD* = 6.16) in one Grade 1 and six Grade 2 classes in three Hebrew language secular schools in a large city in Israel. All schools were defined by the Israeli Ministry of Education as serving similar socially diverse populations with a mix of about 60% students of middle to high SES and about 40% of low SES. Parents were sent a description of the study and asked whether they agreed (a) that their child participate in the study and (b) that schools release their phone number so that we could also invite mothers to participate. Data for children were collected in their schools; the procedure was readministered in two classes after 11 months to 20 boys and 19 girls (mean age = 9 years 1 month). Of the 140 mothers who agreed that we could contact them, 49 refused to participate, repeatedly canceled meetings, or moved to other towns. Data from the other 91 mothers (45 mothers of boys, 46 mothers of girls, mean age = 38 years 1 month) were collected in home visits. Mothers provided background information for the 91 children in the mother–child sample; 84 were living in two-parent families. Schools provided information about maternal education and child age and birth order for the other 72 children. The child and mother–child samples, broken down by child gender and low (12 or fewer years) versus high (at least some higher education) levels of maternal education, are presented in Table 2. Sample comparability was maintained also when we compared the proportions of first- versus later-born children.

Procedure

Data were collected by five female students who were unaware of the study's aims and hypotheses. Children told stories about pictures during school hours, as described for Study 1. Data from mothers were collected in their homes up to 9 months later by a student who had not collected data from that mother's child. The student explained that we had asked children to make up stories about mothers and children and were also interested in how mothers describe things that happen between mothers and children. Mothers were randomly assigned to either a description or a story condition. Each mother in the description condition was asked to recall recent situations in which she and the target child were present and she was angry, she was sad, the child was sad, the mother was sick, and the child was angry. For the neutral cue, she was asked to recall a recent event when mother and child were together. Mothers were encouraged to describe each situation in detail: the setting and background; what each person did, felt, and said; and how it ended. Mothers in the story condition were asked to tell stories about the pictures

Table 2
Participants in Child and Mother–Child Samples of Study 2, by Maternal Education and Gender

	Maternal education			
	High		Low	
	Boys	Girls	Boys	Girls
Child sample	45 (28)	51 (31)	38 (23)	29 (18)
Mother–child sample	26 (29)	29 (32)	19 (21)	17 (19)

Note. Percentages of each sample are in parentheses.

given to children. They were told that each picture showed a mother and her 8 (or 7)-year-old daughter or son, depending on the age and sex of the participating child. Five mothers in this condition said that they did not like to make up stories and were asked to describe events instead. Overall, 49 mothers provided descriptions and 42 told stories; 19 of the latter were also asked to describe actual events. Students had no difficulty writing down mothers' narratives verbatim. Finally, mothers provided information about parents' education and marital status and any medical, psychological, economic, or other problems in the family or the target child.

Scoring

Each measure was scored by different pairs of judges who were all unfamiliar with the hypotheses and did not participate in data collection. Members of each pair independently scored a number of narratives to establish reliability and then divided remaining narratives between them. Different pairs scored stories and descriptions for mothers who provided both.

Active concern and positive emotion. Each story or description about negative emotions was scored for the presence of active concern by each partner, as described for Study 1. Agreement coefficients were calculated separately for CAC and MAC after two pairs of judges scored 50 stories about each situation from child or mother protocols, respectively. Agreement ranged from $\kappa = .85$ for mothers' descriptions of MAC in the mother angry situation to $\kappa = .96$ for both children's and mothers' descriptions of CAC in the mother sick situation. Total CAC and MAC scores over all five stories were highly correlated between scorers, $r(50) = .92$ and $r(50) = .94$ for children, and $r(50) = .92$ and $r(50) = .93$ for mothers. Two new scorers rated narratives about the neutral cue for positive emotional climate with high agreement, $r(50) = .94$ for children and $r(50) = .93$ for mothers.

Patterns of CR. Protocols were assigned to one of four categories of CR or to a fifth, unclear category, as described for Study 1. Agreement coefficients for 40 randomly selected child protocols and 40 randomly selected maternal protocols were $\kappa = .92$ and $\kappa = .90$, respectively.

Child committed compliance. The first author identified stories or descriptions as relevant for scoring if the mother described attempts to influence the child by making suggestions, giving advice, or directing the child to behave in a particular way. Following Kochanska (1997), two scorers rated the 234 relevant narratives for child compliance on a 3-point scale, where a score of 3 reflected eager and committed compliance or cooperation, a score of 2 reflected compliance after pressure or repeated attempts, and a score of 1 reflected noncompliance; stories that did not describe the child's response were coded 0. We could not assess internal reliability because in most protocols at least one story was irrelevant or did not describe the child's response. Final scores for committed compliance were computed as the mean over all relevant stories and were highly correlated between scorers, $r(91) = .89$, $p < .001$.

Results and Discussion

Results are reported in three sections that examine (a) effects of gender and maternal education on representations of concern for the 163 children, (b) effects for the 91 mothers, and (c) relations among representations of concern and compliance in the 91 dyads.

Child Data

Representations of active concern and patterns of CR. We predicted that the effects of child gender on representations of CAC and patterns of CR reported in Study 1 would be replicated at both high and low levels of maternal education. As expected, a 2 (gender) × 2 (maternal education) analysis of variance (ANOVA) for children's representations of child concern yielded only a significant main effect for gender, $F(1, 159) = 12.24$, $p < .001$ ($M_{\text{girls}} = 1.97$, $SD = 1.30$ and $M_{\text{boys}} = 1.25$, $SD = 1.34$). In contrast, a 2 (gender) × 2 (maternal education) ANOVA for representations of maternal concern yielded unanticipated main effects for gender $F(1, 159) = 5.13$, $p < .05$, and maternal education, $F(1, 159) = 5.65$, $p < .05$, which were qualified by a significant interaction effect, $F(1, 159) = 7.16$, $p < .01$, $MSE = 1.51$. Follow-up analyses indicated that these effects derived from the low levels of maternal concern described by boys in the low maternal education group (see Table 3 for cell means). Girls described significantly more instances of maternal concern than did boys in the low, $F(1, 159) = 10.32$, $p < .001$, $MSE = 1.51$, but not the high maternal education group. In addition, representations of maternal concern were significantly more frequent at high than at low levels of maternal education among boys, $F(1, 159) = 13.53$, $p < .001$, $MSE = 1.51$, but not girls.

The results for patterns of CR were also more complex than anticipated because a 5 (pattern) × 2 (gender) × 2 (maternal education) loglinear analysis

Table 3

Means for Children's and Mothers' Representations by Maternal Education and Gender, Study 2

	High maternal education				Low maternal education			
	Boys		Girls		Boys		Girls	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Children								
CAC	1.36	1.42	1.94	1.27	1.13	1.26	2.03	1.38
MAC	2.02	1.27	1.94	1.33	1.03	1.02	2.00	1.23
Positive emotion	3.36	1.33	3.55	1.39	2.61	1.29	3.52	1.27
n	45		51		38		29	
Mothers								
CAC	1.27	0.87	1.83	1.04	0.84	0.90	2.00	1.00
MAC	2.08	0.94	2.10	0.90	1.53	0.96	2.06	1.20
Positive emotion	3.85	1.05	3.93	1.03	3.53	1.12	4.00	1.23
Child willing compliance	2.38	0.64	2.49	0.65	2.01	0.69	2.57	0.71
n	26		29		19		17	

Note. CAC = child active concern; MAC = maternal active concern.

yielded a significant third-order effect, $\chi^2(4) = 8.68$, $p < .05$ (see Table 4 for cell frequencies). Follow-up 2 × 2 × 2 loglinear analyses for each pattern, with the target pattern coded 1 and all other patterns coded 0 contributed to clarifying this effect. As predicted, the mutual CR pattern was significantly associated only with gender, $\chi^2(1) = 18.09$, $p < .001$, and was described far more frequently by girls (49%) than by boys (16%). In addition, the tendency noted in Study 1 for more boys than girls to describe a relationship in which only the child displayed active concern was significant for the larger sample of Study 2, $\chi^2(1) = 8.20$, $p < .01$; 16% of boys' protocols but only 3% of girls' protocols were scored as child CR. The third-order effect for Pattern × Gender × Maternal Education was significant for both the maternal CR, $\chi^2(1) = 4.92$, $p < .05$, and the nonresponsive pattern, $\chi^2(1) = 3.97$, $p < .05$. Further analyses and the frequencies presented in Table 4 showed that results for the high maternal education group replicated those of Study 1 because more boys than girls described a pattern of maternal CR, $\chi^2(1) = 6.25$, $p < .05$, and gender did not affect the frequency of the nonresponsive pattern. In contrast, in the low maternal education group, boys were no more likely than girls to describe a pattern of maternal CR but were more likely to describe the nonresponsive pattern, $\chi^2(1) = 4.63$, $p < .05$.

Table 4

Percentages of Children's Protocols Assigned to Each Pattern of Mother–Child CR by Maternal Education and Gender, Study 2

	High maternal education		Low maternal education		
	Boys	Girls	Boys	Girls	All
Mutual CR	18	47	11	41	29
Maternal CR	36	14	16	24	22
Child CR	13	4	16	0	9
No CR	24	31	50	24	32
Unclear	9	4	8	10	8
n	45	51	38	29	163

Note. CR = concerned responsiveness.

Emotional climate in the bench story. A 2 (gender) × 2 (maternal education) ANOVA yielded an unanticipated effect for gender, $F(1, 159) = 6.53$, $p < .05$ ($M_{\text{girls}} = 3.54$, $SD = 1.34$ and $M_{\text{boys}} = 3.01$, $SD = 1.36$) and a marginally significant effect for maternal education ($p < .10$). The means presented in Table 3 indicate that both effects derived mainly from the low level of positive emotion described by boys in the low maternal education group. Thus, to summarize, results for the high maternal education group replicated those for the similar sample of Study 1, and results for girls were similar in both social groups. In contrast, the boys in the low maternal education group described lower levels than did the girls not only of child concern but also of maternal concern and positive emotion. Low-SES boys were also less likely than were boys in the high maternal education group to describe a pattern of maternal CR and were more likely to describe a relationship in which neither partner displayed active concern.

Stability of children's representations. Measures for children who were retested after 11 months were quite stable, $r(39) = .45$, $p < .01$ for CAC, and $r(39) = .58$, $p < .001$ for MAC. Protocols were assigned to the same category of CR in 79% of cases.

Mother Data

Representations of active concern and patterns of CR. Mean levels of CAC and MAC did not differ significantly in descriptions versus stories, and these procedures yielded similar distributions of patterns of CR (33% mutual, 27% maternal, 4% child, 33% nonresponsive, and 4% unclear for descriptions and 33%, 19%, 2%, 10%, respectively, for stories). Protocols of mothers who provided descriptions and stories were assigned to the same category of CR in 90% of cases. These results allayed our concern that descriptions might be more positively biased than stories and justified combining data from both procedures. We analyzed descriptions for mothers who also told stories.

Overall, the results for mothers were consistent with the initial predictions. A 2 (gender) × 2 (maternal education) ANOVA for maternal descriptions of CAC yielded only the predicted significant main effect for gender, $F(1, 87) = 17.42$, $p < .001$; mothers described more instances of active concern by daughters ($M = 1.84$, $SD = 0.98$) than by sons ($M = 1.09$, $SD = 0.90$). The analysis for representations of maternal concern did not yield any significant effects involving gender or maternal education, but the means presented in Table 3 show that low-SES mothers of sons tended to describe fewest instances of both MAC and CAC.

Few maternal protocols were scored as child CR or unclear (see Table 5 for cell frequencies). Loglinear modeling of associations between gender, maternal education, and the three other patterns ($n = 82$) retained only the predicted association between patterns and gender, $\chi^2(2) = 10.58$, $p < .01$. Follow-up analyses confirmed that more mothers of daughters (54%) than sons (20%) described a relationship of mutual CR, $\chi^2(1) = 10.30$, $p < .001$, but the predicted association between gender and the maternal CR pattern was only marginally significant ($p < .10$). As was the case also for children, this pattern was more frequent in mother–son than in mother–daughter protocols in the high maternal education group, but in the low maternal education group, mothers of boys were no more likely than mothers of girls to describe a pattern of maternal CR and were more likely to describe a nonresponsive relationship (see Table 5 for cell frequencies).

Positive emotional climate and child compliance. Positive emotion did not differ significantly by gender or maternal education, but a 2 (gender) × 2 (maternal education) ANOVA for maternal descriptions of child compliance yielded a significant main effect for gender, $F(1, 87) = 4.17$, $p < .05$. Overall, mothers described girls as more cooperative and compliant ($M = 2.53$, $SD = 0.66$) than boys ($M = 2.26$, $SD = 0.69$), but the means presented in Table 3 indicate that this difference was marked only in the low maternal education group.

Relations Among Representations in Mother–Child Dyads and Narratives

Patterns of CR within mother–child dyads are presented in Table 6. Although data for mothers and children were collected in different arenas, with a time lag of up to 9 months, there was considerable congruence between their representations. Thus, for each of the three relatively frequent patterns of mutual, maternal, and no CR, mother and child protocols were assigned to the same category in over 70% of dyads. Low expected cell frequencies precluded statistical analysis, but the correlations presented in Table 7 confirm that equivalent measures derived from children's and mothers' narratives were significantly correlated in both mother–son and mother–daughter dyads. This was the case both for descriptions of CAC and MAC and for descriptions of positive emotional climate; no correlations differed significantly at high versus low levels of maternal education.

The correlations presented in Table 7 also confirmed that descriptions of maternal and child concern were significantly correlated in mother–daughter but not mother–son narratives and dyads; the difference between correlations was significant for children's narratives, $z = 2.22$, $p < .05$. In keeping with the results of Study 1, positive emotional climate was significantly correlated with active concern in mother–daughter but not mother–son narratives

Table 5

Percentages of Maternal Protocols Assigned to Each Pattern of Mother–Child CR by Maternal Education and Child Gender, Study 2

	High maternal education		Low maternal education		
	Boys	Girls	Boys	Girls	All
Mutual CR	19	52	16	41	33
Maternal CR	39	14	21	18	23
Child CR	4	3	0	6	3
No CR	35	28	53	24	34
Unclear	4	3	11	12	6
n	26	29	19	17	91

Note. CR = concerned responsiveness.

Table 6

Cross-Tabulation of Patterns of CR Within Mother–Child Pairs, Study 2

	Children				
	Mutual	Maternal	Child	Nonresponsive	Unclear
Mothers					
Mutual	82	9	14	3	50
Maternal	7	73	14	7	0
Child	4	0	14	3	0
Nonresponsive	7	18	29	82	0
Unclear	0	0	29	7	50
n	28	22	7	28	6

Note. Data presented as percentages of child protocols assigned to each category of CR. CR = concerned responsiveness.

Table 7
Correlations Between Measures for Children and Mothers in 91 Mother – Child Pairs

	CMAC		CCAC		MMAC		MCAC		CPE		MPE	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
CCAC	.25	.63***										
MMAC	.38**	.32*	.11	.25								
MCAC	.07	.36*	.31*	.57***	.11	.36*						
CPE	.34*	.36**	.14	.55***	.30*	.22	.07	.39**				
MPE	.31*	.25	.12	.32*	.63***	.49***	.11	.30*	.33*	.44**		
MCCP	.53***	.49***	.32*	.31*	.65***	.40***	.32*	.49***	.39**	.38**	.46**	.34*

Note. The first letter denotes whether the measure was derived from child or maternal narratives. CAC = child active concern; MAC = maternal active concern; PE = positive emotion; CCP = child committed compliance.
$*p < .05$. $**p < .01$. $***p < .001$.

and dyads. In contrast, maternal concern was similarly and usually significantly correlated with descriptions of both child compliance and positive emotional climate in all narratives and dyads. Despite the different correlates of child concern and child committed compliance, these kinds of responsiveness were significantly correlated for both girls and boys.

Supplementary Analyses

Representations of active concern were not significantly associated with birth order, number of siblings, or maternal reports of problems in the family or child, and the number of problems reported did not differ by gender or maternal education. Although girls might be more interested than boys in socioemotional themes, gender was not related to the word count of children's stories, and this measure of verbal fluency was not significantly correlated with either MAC, $r(163) = .05$ or CAC, $r(163) = .11$.

General Discussion

Results from two studies provided novel evidence of gender differences in mother – child reciprocal responsiveness and suggested that these were due in large part to the differential contributions of girls and boys to the interactive arena. Children and mothers described girls as acting more often than boys to alleviate maternal distress and to maintain or restore relational harmony. In contrast, with the exception of children in the low-SES sample of Study 2, participants did not describe mothers as showing more concern toward daughters than sons. Thus, far more mother – daughter than mother – son narratives described a relationship of "mutual CR" reminiscent of the reciprocal care and concern that has emerged in studies of close female friendships (Burke & Fuqua, 1987; Wright, 1982).

We examined representations rather than behaviors, but studies have shown consistent relations between narrative and observational measures of attachment and maternal sensitivity (Bar-Haim, Sutton, Fox, & Marvin, 2000; Bretherton, Ridgeway, & Cassidy, 1990; Solomonica-Levi et al., 2001) and between child narrative, mother report, and behavioral measures of children's concern for others (Davidov & Grusec, 2006). The results of Study 2 confirmed further that child representations of active concern and patterns of CR were consistently associated with equivalent measures derived up to 9 months later from their mothers, most of whom provided detailed descriptions of actual interactions with the child. Thus, we can be fairly confident that the narratives of both children and mothers described important aspects of their experiences and behaviors in the mother – child relationship.

Many studies have shown that females tend to show greater interpersonal care and responsibility than do males, as peers, friends, spouses, and parents (Maccoby, 1998) and also in early interactions with the mother (Kiang et al., 2004; Kochanska et al., 1999; Radke-Yarrow et al., 1994; Zahn-Waxler et al., 1992). Interestingly, however, prior studies have not examined the implications of these differences for the development of reciprocity in mother – daughter and mother – son dyads. Rather, analyses of mother – child mutual responsiveness have emphasized the role of maternal sensitive concern in creating a relational system in which children of both sexes are motivated to cooperate and comply with maternal initiatives, and influence, and engage with the mother in mutually enjoyable interactions (Biringen & Robinson, 1991; Kochanska, 1997, 2002; Maccoby & Martin, 1983). Against this background, our main contribution was in conceptualizing active concern as a feature not only of the mother but also of the child and of the interactions between them. This approach yielded a novel framework for characterizing the mother – child relationship in terms of patterns of CR and provided the clearest empirical evidence to date that mother – daughter dyads are more likely than mother – son dyads to develop a relationship of mutual care and concern already in childhood (Chodorow, 1978; Jordan & Surrey, 1986).

These results are consistent with our proposal that other researchers did not find gender differences in mother – child reciprocity because they examined different kinds of child and mutual responsiveness. Representations of child concern and committed compliance were moderately correlated, however, and girls tended to score higher than boys also on committed compliance. This was the case also in studies of mother – child mutually responsive orientation (e.g., Kochanska, 1997; Kochanska, Coy, & Murray, 2001). Further studies might address the intriguing possibility that Kochanska's strategy of assessing mutually responsive orientation as an aggregate of child committed compliance, maternal sensitivity, and shared positive affect might mask possible differential contributions of girls and boys to the development of this kind of mother – child reciprocity as well. However, although we emphasized differences in the level of concern that girls and boys bring to interactions with the mother, our results also revealed another way in which gender seemed to influence mutual concern more than it does other kinds of mother – child reciprocity. In keeping with prior findings (see Kochanska, 2002; Maccoby, 1998, for reviews), associations between maternal responsiveness, child willing compliance, and positive affect were similar and significant across all narratives and dyads. Therefore, boys did not fail to respond to maternal concern, and sensitive parenting seemed to orient children of both sexes to interact positively and cooperatively with sensitive mothers. In contrast, representations of maternal and child concern were significantly associated only in mother – daughter narratives and dyads. Thus, mutual responsiveness in mother – daughter and mother – son interactions differed not only because girls showed more active concern, and to some extent compliance than boys, but also because girls seemed more likely to reciprocate maternal concern in kind.

There is some prior evidence that links between parenting and children's empathy were more consistent in mother – daughter than in mother – son dyads (Eisenberg & Fabes, 1998; Hastings, Utendale, & Sullivan, 2007; Zahn-Waxler, Cole, & Barrett, 1991; but see Kiang et al., 2004, for a recent exception). Our data cannot explain why this might be the case, but one possibility is that mothers induct daughters more than sons into a relationship of reciprocal concern. Responsive mothering is crucial in creating a system in which children are receptive to maternal influence, but relations between maternal sensitivity and child outcomes are moderated by parenting goals and practices (Grusec & Goodnow, 1994; Hastings & Rubin, 1999), some of which may differ for boys and girls. Specifically, mothers seem to value and strive to socialize willing compliance in both sons and daughters (Hastings & Grusec, 1998), but gender stereotypes, mother's prior social histories, and early gender differences in sociability might converge in orienting mothers to value, expect, and socialize active concern for the mother in girls more than in boys (see also Brody, 1999). If mothers expect daughters to show more concern than sons, this could also explain why maternal descriptions of mother – child emotionally positive interactions were significantly correlated with their descriptions of girls' but not boys' active concern. Few studies have directly examined maternal socialization of care and concern, but mothers do seem to talk to and discipline young children in ways that direct girls more than boys to consider other people's perspectives and feelings and to assume responsibility for alleviating distress and repairing relationships (Brody, 1999; Fivush, 1993; Leaper et al., 1998; Smetana, 1989). Any such divergence in reciprocal concern in mother – daughter as compared with mother – son dyads should then increase as gender differences in relational styles are reinforced in the same-sex peer group (Maccoby, 1990).

Although the school-age children in our samples already had considerable peer experience, other scholars have also proposed that the mother – child relationship plays a role in orienting girls to display care and concern in relationships and orienting boys to cooperate with, but not necessarily reciprocate concern in their interactions with, females (Brody, 1999; Robinson, Little, & Biringen, 1993). Indeed, the results for the high-SES samples confirmed that more mother – son than mother – daughter narratives described a pattern of maternal CR reminiscent of the unequal division of support work that has emerged in studies of cross-sex relationships in later life (Fishman, 1983; Tannen, 1993). Interestingly, boys were also more likely than girls to describe the other nonreciprocal pattern, in which only the child showed active concern. The intriguing implication that child concern develops in response to maternal concern in girls, but may develop in response to maternal deficits in at least some boys, must be treated with caution because, overall, both boys and girls who described unconcerned mothers also described unconcerned children. In support, however, Radke-Yarrow et al. (1994) found that maternal depression was necessary to elicit caring behavior in sons but undermined care for the mother in girls.

Finally, we must address the unanticipated interactions between gender and SES, which we assessed in terms of high versus low levels of maternal education. Descriptions of concern and compliance in mother – daughter interactions did not differ at higher and lower levels of SES, but a striking 50% of low-SES boys told stories in which neither mother nor child showed concern for the other or the relationship. Mothers' narratives differed significantly only by gender, but low-SES mothers of sons also tended to describe low levels of maternal and child concern and were most likely to describe a nonresponsive relationship. Descriptions of emotional climate and child committed compliance were also most negative in this group. A close reading of maternal protocols indicated further that low-SES mothers of sons were most likely to spontaneously complain that the child did not listen to her and was hard to talk to.

If these sons were relatively uncommunicative and resistant to maternal influence, they might also have evoked less sensitive, more power-assertive parenting, which tends to further undermine children's responsiveness to gentler forms of influence to create an escalating cycle of aversive interactions (Campbell, Pierce, Moore, & Marakovitz, 1996; Sanders, Dadds, Johnston, & Cash, 1992). Conversely, low-SES mothers and daughters, at least in our sample, may have maintained close and positive relations at least in part because they talked more. In their comprehensive study of British working-class and middle-class mothers and their 7-year-old sons or daughters, Newson and Newson (1976) found somewhat similar patterns whereby mothers and daughters chatted to and confided in one another in both social groups, middle-class sons talked to their mothers but were less likely than girls to invite and receive maternal confidences, and mother – child interactions were most distant and aversive in working-class mother – son dyads. Of particular interest, working-class mothers attributed these negative interactions mainly to the growing influence of fathers and male peers whom they saw as encouraging their sons to resist and disparage maternal influence and companionship.

We must be cautious in generalizing from our results because of the small sample, the sometimes insignificant trends, and the mixed results of some other studies. For example, maternal sensitivity tends to be lower in low-SES samples than in high-SES samples (see Conger & Dogan, 2007; Hoff, Laursen, & Tardiff, 2002, for reviews), but in our Study 2, this was the case only for mothers of boys. In addition, as we reviewed in an earlier section, studies that targeted low-SES samples have yielded mixed findings as to whether mothers behaved more sensitively toward sons or daughters, and Eisenberg (1996) reported more aversive conflicts in low-SES mother – daughter than mother – son dyads. Moreover, although our discussion focused on unresponsive low-SES mother – son dyads, it is important to emphasize that about a quarter of the narratives in other groups also described unresponsive and usually aversive interactions. Finally, reviews have identified both SES (Hoff et al., 2002) and gender (Leaper & Smith, 2004; Leaper et al., 1998) influences on mother – child talk, but few studies have examined both sources of influence. Thus, further research is necessary to examine our tentative interpretation that low-SES mother – son dyads displayed such low levels of concern at least in part because boys growing up in traditional working-class cultures might be least likely to listen to their mothers' "voice."

Before we address some additional implications, it is important to note several limitations of our studies. First, the narratives provided rich and coherent information as to how representations of concern and other aspects of mother – child interactions were related within individual protocols and mother – child pairs, but it is important to validate our framework and results for additional, preferably behavioral, measures of concern and compliance. Second, the results for high-SES samples provided consistent support for our proposal that there are gender differences in child, more than in maternal concern, but one must be cautious in drawing conclusions from null results. Third, we did not consider other arenas that might influence school-age children's CR in the mother–child relationship. Peers and fathers may be more influential than mothers in sex-role socialization and may play a particularly important role in fostering an orientation of care and concern in girls and orienting boys to construct a sense of self based in part on the repression of such feminine qualities (Gilligan, 1996; Maccoby, 1998; Ruble & Martin, 1998; Siegal, 1987). Fourth, our interpretation that any such influences might be particularly marked in more traditional working-class cultures must be treated with caution in view of the small sample and the mixed results of some other studies. Longitudinal, multimethod studies that would examine the development and correlates of child and reciprocal concern between early and middle childhood in different social groups would contribute to addressing these limitations.

Focusing on levels and patterns of coordination between maternal and CAC shed new light on the contribution of the child to reciprocal responsiveness in mother–son and mother–daughter dyads and revealed some intriguing continuities in the ways in which boys and girls behave with the mother and in later relationships with females. Our concepts and results also suggest some interesting directions for further research. Distinguishing between global maternal sensitivity and specific practices more directly relevant to the induction of concern might reveal more consistent linkages than hitherto between parenting and the development of empathic concern in both boys and girls (see also Davidov & Grusec, 2006; Eisenberg, Fabes, Carlo, Speer, & Switzer, 1993). Our framework and methods might also contribute to understanding the joint influences of both child and parent gender. For example, gender differences in concern imply that fathers will be less likely than mothers to develop a relationship of mutual CR, especially with sons. In support, Maccoby (1998) reported that mother–daughter pairs were most successful and father–son dyads were least successful on a task of referential communication because of gender differences in the referential skills of both children and parents. Our narrative procedures might also be well suited, however, to reveal other ways in which fathers might promote mutually binding relations with their children.

Our distinctions between different patterns of CR might also contribute to reconciling approaches that consider empathic concern for the mother to be an important component of prosocial behavior (Hastings et al., 2000; Kiang et al., 2004) with approaches that consider child concern for the mother to be rather maladaptive (Freud, 1946; Winnicott, 1965). This might be the case if concern develops in response to poor mothering or is expressed in age-inappropriate ways. However, most stories scored for CAC described children who cared about, rather than taking care of the mother, and most narratives that described concerned children also described concerned parents. Thus, the role of children's concern in their social and emotional adjustment should depend mainly on the degree to which children also experience the mother as concerned and sensitive. Further studies might, however, address the possibility that as long as interpersonal care and concern are viewed mainly as "women's work," high levels of concern for the mother might be more problematic for boys than for girls. Thus, such concern might be more likely to develop in response to maternal deficits in boys (Radke-Yarrow et al., 1994) and is less congruent with the relational style of the male than the female peer group (Maccoby, 1998; Maltz & Borker, 1982). Last, but not least, our results highlight the importance of examining gender patterns in the mother–child relationship in different social and cultural groups.

References

Ainsworth, M. S., Blehar, M. C., Waters, E., & Wall, S. (1978). *Patterns of attachment: A psychological study of the strange situation*. Hillsdale, NJ: Erlbaum.

Bar-Haim, Y., Sutton, D. B., Fox, N. A., & Marvin, R. S. (2000). Stability and change of attachment at 14, 24, and 58 months of age: Behavior, representation, and life events. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *41*, 381–388.

Barnett, D., Kidwell, S. L., & Leung, K. H. (1998). Parenting and preschooler attachment among low-income African-American families. *Child Development*, *69*, 1657–1671.

Bartky, S. L. (1990). Feeding egos and tending wounds: Deference and disaffection in women's emotional labor. In S. L. Bartky, *Femininity and domination: Studies in the phenomenology of oppression* (pp. 99–119). New York: Routledge.

Bell, S. M., & Ainsworth, M. D. (1972). Infant crying and maternal responsiveness. *Child Development*, *43*, 1171–1190.

Biringen, Z., Brown, D., Donaldson, L., Green, S., Krcmarik, S., & Lovas, G. (2000). Adult Attachment Interview: Linkages with dimensions of emotional availability for mothers and their pre-kindergarteners. *Attachment and Human Development*, *2*, 188–202.

Biringen, Z., Emde, R. N., Brown, D., Lowe, L., Myers, S., & Nelson, D. (1999). Emotional availability and emotional communication in naturalistic mother-child interactions: Evidence for gender relations. *Journal of Social Behavior and Personality, 14*, 463–478.

Biringen, Z., & Robinson, J. (1991). Emotional availability in mother-child interactions: A reconceptualization for research. *American Journal of Orthopsychiatry, 61*, 258–271.

Bretherton, I., Ridgeway, D., & Cassidy, J. (1990). Assessing internal working models of the attachment relationship: An attachment story completion task for 3-year olds. In M. T. Greenberg & D. Ciccetti (Eds.), *Attachment in the preschool years: theory, research and intervention* (pp. 273–308). Chicago: University of Chicago Press.

Brody, L. R. (1999). *Gender, emotion, and the family.* Cambridge, MA: Harvard University Press.

Burke, R. A., & Fuqua, D. R. (1987). Sex differences in same and cross-sex supportive relationships. *Sex Roles, 17*, 339–352.

Campbell, S. B., Pierce, E. W., Moore, G., & Marakovitz, S. (1996). Boys' externalizing problems at elementary school age: Pathways form early behavior problems, maternal control and family stress. *Development and Psychopathology, 8*, 701–719.

Chodorow, N. (1978). *The reproduction of mothering: Psychoanalysis and the sociology of gender.* Berkeley, CA: University of California Press.

Conger, R. D., & Dogan, S. J. (2007). Social class and socialization in families. In J. E. Grusec & P. D. Hastings (Eds.), *Handbook of socialization* (pp. 433–461). New York: Guilford.

Davidov, M., & Grusec, J. E. (2006). Untangling the links of parental responsiveness to distress and warmth to child outcomes. *Child Development, 77*, 44–58.

Dix, T. (1991). The affective organization of parenting: Adaptive and maladaptive processes. *Psychological Bulletin, 110*, 3–25.

Dodge, K. A., Pettit, G. S., & Bares, J. E. (1994). Socialization mediators of the relation between socioeconomic status and child conduct problems. *Child Development, 65*, 649–665.

Eberly, M. B., & Montemayor, R. (1998). Doing good deeds: An examination of adolescent prosocial behavior in the context of parent-adolescent relationships. *Journal of Adolescent Research, 13*, 403–432.

Eisenberg, A. R., (1996). The conflict talk of mothers and children: Patterns related to culture, SES, and gender of child. *Merrill-Palmer Quarterly, 42*, 438–458.

Eisenberg, N., & Fabes, R. A. (1998). Prosocial development. In W. Damon & N. Eisenberg (Eds.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (5th ed., pp. 702–778). New York: Wiley.

Eisenberg, N., Fabes, R. A., Carlo, G., Speer, A., & Switzer, G., D. (1993). The relations of empathy-related emotions and maternal practices to children's comforting behaviors. *Journal of Experimental Child Psychology, 55*, 131–150.

Fishman, P. M. (1983). Interaction: The work women do. In B. Thorne, C. Kramerae, & N. Henley (Eds.), *Language, gender and society* (pp. 89–101). Cambridge, MA: Newbury House.

Fivush, R. (1993). Emotional content of parent-child conversations about the past. In C. A. Nelson (Ed.), *Memory and affect in development. Minnesota Symposium on Child Psychology* (Vol. 26, pp. 39–78). Minneapolis, MN: University of Minnesota Press.

Freud, A. (1946). *The ego and the mechanisms of defense.* New York: International Universities Press.

Gilligan, C. (1996). The centrality of relationship in human development: A puzzle, some evidence, and a theory. In. G. G. Noam & K. W. Fischer (Eds.), *Development and vulnerability in close relationships* (pp. 237–261). Hillsdale, NJ: Erlbaum.

Grusec, J. E., & Goodnow, J. J. (1994). Impact of parental discipline methods on the child's internalization of values: A reconceptualization of current points of view. *Developmental Psychology, 32*, 99–1007.

Hastings, P. D., & Grusec, J. E. (1998). Parenting goals as organizers of responses to parent-child disagreement. *Developmental Psychology, 34*, 465–479.

Hastings, P. D., & Rubin, K. H. (1999). Predicting mothers' beliefs about preschool-aged children's social behavior: Evidence for maternal attitudes moderating child effects. *Child Development, 70*, 722–741.

Hastings, P. D., Utendale, W. T., & Sullivan, C. (2007). The socialization of prosocial development. In J. E. Grusec & P. D. Hastings (Eds.), *Handbook of socialization* (pp. 638–664). New York: Guilford.

Hastings, P. D., Zahn-Waxler, C., Robinson, J., Usher, B., & Bridges, D. (2000). The development of concern for others in children with behavior problems. *Developmental Psychology, 36*, 531–546.

Hoff, E., Laursen, B., & Tardiff, T. (2002). Socioeconomic status and parenting. In M. H. Bornstein (Ed.), *Handbook of parenting: Vol. 2: Biology and ecology of parenting* (pp. 231–252). Mahwah, NJ: Erlbaum.

Jordan, J. V., & Surrey, J. L. (1986). The self-in-relation: Empathy and the mother-daughter relationship. In T. Bernay & D. W. Cantor (Eds.), *The psychology of today's woman: New psychoanalytic visions* (pp. 81–104). Cambridge, MA: Harvard University Press.

Kiang, L., Moreno, A. J., & Robinson, J. J. (2004). Maternal preconceptions about parenting predict child temperament, maternal sensitivity, and children's empathy. *Developmental Psychology, 40*, 1081–1092.

Kochanska, G. (1997). Mutually responsive orientation between mothers and their young children: Implications for early socialization. *Child Development, 68*, 94–112.

Kochanska, G. (2002). Mutually responsive orientation between mothers and their young children: A context for the early development of conscience. *Current Directions in Psychological Science, 11*, 191–195.

Kochanska, G., Coy, K. C., & Murray, K. T. (2001). The development of self-regulation in the first four years of life. *Child Development, 68*, 94–112.

Kochanska, G., & Forman, D. R., & Coy, K. C. (1999). Implications of the mother-child relationship in infancy: Socialization in the second year of life. *Infant Behavior and Development, 22*, 249–265.

LaFreniere, P., Masataka, N., Butovskaya, M., Chen, Q., Dessen, M. A., Alwanger, K, et al. 2002). Cross-cultural analysis of social competence and behavior problems in preschoolers. *Early Education and Development, 13,* 201–219.

Leaper, C. (2002). Parenting girls and boys. In M. H. Bornstein (Ed.), *Handbook of parenting: Vol. 1: Children and parenting* (pp. 189–225). Mahwah, NJ: Erlbaum.

Leaper, C., Anderson, A., K., & Sanders, P. (1998). Moderators of gender effects on parents' talk to children: A meta-analysis. *Developmental Psychology, 34,* 3–27.

Leaper, C., & Smith, T. E. (2004). A meta-analytic review of gender variations in children's language use: Talkativeness, affiliative speech, and assertive speech. *Developmental Psychology, 40,* 993–1027.

Lytton, H., & Romney, D. M. (1991). Parents' differential socialization of boys and girls: A meta-analysis. *Psychological Bulletin, 109,* 267–296.

Maccoby, E. (1990). Gender and relationships: A developmental account. *American Psychologist, 45,* 513–520.

Maccoby, E. E. (1998). *The two sexes: Growing up apart, coming together.* Cambridge, MA: Harvard University Press.

Maccoby, E. E., & Martin, J. A. (1983). Socialization in the context of the family: Parent-child interactions. In P. H. Mussen & E. M. Hetherington (Eds.), *Handbook of child psychology* (4th ed., Vol. 4, pp. 1–102). New York: Wiley.

Maccoby, E. E., & Parpal, M. (1985). Maternal responsiveness and subsequent child compliance. *Child Development, 56,* 1326–1334.

Maltz, D. N., & Borker, R. A. (1982). *A cultural approach to male-female miscommunication.* In J. J. Gumperz (Ed.), Language and social identity (pp. 195–216). Cambridge, England: Cambridge University Press.

Newson, J., & Newson, G. (1976). *Seven years old in the home environment.* Oxford, England: Wiley.

Oppenheim, D., Emde, R. N., & Warren, S. (1997). Children's narrative representations of mothers: Their development and associations with child and mother adaptation. *Child Development, 68,* 127–138.

Radke-Yarrow, M., Zahn-Waxler, C., Richardson, N. T., Susman, A., & Martinez, P. (1994). Caring behavior in children of clinically depressed and well mothers. *Child Development, 65,* 1405–1414.

Robinson, J., Little, C., & Biringen, Z. (1993). Emotional communication in mother-toddler relationships: Evidence for early gender differentiation. *Merrill-Palmer Quarterly, 39,* 496–517.

Ruble, D. N., & Martin, C. L. (1998). Gender development. In W. Damon & N. Eisenberg (Eds.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (5th ed., pp. 933–1016). New York: Wiley.

Sanders, M. R., Dadds, M. R., Johnston, B. M., & Cash, R. (1992). Childhood depression and conduct disorder: Behavioral, affective and cognitive aspects. *Journal of Abnormal Psychology, 101,* 495–504.

Siegal, M. (1987). Are sons and daughters treated more differently by fathers than by mothers? *Developmental Review, 7,* 183–209.

Smetana, J. G., (1989). Toddlers' social interactions in the context of moral and conventional transgressions in the home. *Developmental Psychology, 25,* 499–509.

Solomonica-Levi, D., Yirmiya, N., Erel, O., Samet, I., & Oppenheim, D. (2001). The associations between observed maternal behavior, children's narrative representations of mothers, and children's behavior problems. *Journal of Social and Personal Relationships, 18,* 673–690.

Tannen, D. (1993). *Gender and conversational interaction.* New York: Oxford University Press.

Whiting, B. B., & Edwards, C. P. (1988). *Children of different worlds: The formation of social behavior.* Cambridge, MA: Harvard University Press.

Winnicott, D. W. (1965). *The maturational process and the facilitating environment.* New York: International Universities Press.

Wright, P. H. (1982). Men's friendships, women's friendships, and the alleged inferiority of the latter. *Sex Roles, 8,* 1–20.

Zahn-Waxler, C., Cole, P., & Barrett, K. C. (1991). Guilt and empathy: Sex differences and implications for the development of depression. In J. Garber & K. Dodge (Eds.), *The development of empathic regulation and dysregulation* (pp. 243–272). New York: Cambridge University Press.

Zahn-Waxler, C., Radke-Yarrow, M., Wagner, E., & Chapman, M. (1992). Development of concern for others. *Developmental Psychology, 28,* 126–136.

Child Development, July/August 2008, Volume 79, Number 4, Pages 852–865

Development of Feedforward Control in a Dynamic Manual Tracking Task

Dominique van Roon, Karen Caeyenberghs,
and Stephan P. Swinnen
Katholieke Universiteit Leuven

Bouwien C. M. Smits-Engelsman
Katholieke Universiteit Leuven, Avans+, University for Professionals, and Radboud University Nijmegen

To examine the development of feedforward control during manual tracking, 117 participants in 5 age groups (6 to 7, 8 to 9, 10 to 11, 12 to 14, and 15 to 17 years) tracked an accelerating dot presented on a monitor by moving an electronic pen on a digitizer. To remain successful at higher target velocities, they had to create a predictive model of the target's motion. The ability to track the target at higher velocities increased, and the application of a feedback-based step-and-hold strategy decreased with age, as shown by increases in maximum target velocity and decreases in number of stops between ages 6–7 and 8–9 and between ages 8–9 and 10–11. The ability to exploit feedforward control in a dynamic tracking task improves significantly with age.

According to von Hofsten (1993, 2007) the development of action is a matter of acquiring prospective control. When interacting with the environment, actions have to become increasingly future oriented. Predictions have to be made about what is going to happen next. With increasing age, these predictions become more and more accurate because the processing of appropriate predictive perceptual information in the environment becomes faster. In addition, children learn the rules and regularities that govern events and are able to cognitively represent such events. Visually or manually tracking fast-moving objects are examples of tasks in which prospective (or feedforward) control is indispensable.

The development of predictive *visual tracking* has been studied extensively in the past (e.g., Bertenthal, Longo, & Kenny, 2007; Gredebäck, von Hofsten, Karlsson, & Aus, 2005; von Hofsten & Rosander, 1997). Even very young infants in the 1st year of their lives appear to be able to predict object motion when visually tracking this object. In addition, they are able to use the past history of a moving object to build an internal representation of the object's motion, which can be used to predict when and where the object will reappear after a temporary occlusion (e.g., Gredebäck & von Hofsten, 2004; Spelke, Katz, Purcell, Ehrlich, & Breinlinger, 1994). Also, when a child wants to grasp a moving object, he or she has to be able to predict its future location and anticipate to that. Infants as young as 8 months appear to be able to catch objects moving at 120 cm/s (von Hofsten, 1983).

Although visual pursuit of a moving target appears to develop very early in life, *manually tracking* a moving target appears to develop later. Although manually tracking a visible target always requires visual tracking, and the term visuomanual tracking is therefore used in many studies, we will use the term manual tracking to prevent confusion. Several developmental studies on manual tracking were performed previously. Some investigators used a rotary pursuit apparatus (e.g., Davol, Hastings, & Klein, 1965; Dunham, Allan, & Winter, 1985), in which a participant had to maintain contact between a predictably rotating target spot and a handheld stylus. In other studies, the task was to track a target light, projected on a vertical screen and moving predictably along a straight line, by moving the splinted forearm in the horizontal plane (e.g., Mounoud, Viviani, Hauert, & Guyon, 1985; Zanone, 1990). Van der Meulen, Gooskens, Willemse, Denier van der Gon, and Gielen (1990) used a horizontal rail along which a handle could be moved. Two arrays of light emitting diodes (LED) (indicating the target, which moved unpredictably, and the handle position) were visible above the track. In our tracking study, a digitizing tablet was used. Participants had to move an electronic pen on this tablet to track a circularly moving target presented on a computer monitor in front of them.

Support for this study was provided through a grant from the Research Programme of the Research Foundation—Flanders (FWO, Levenslijn # 7.0004.05). D. van Roon was funded by a visiting postdoctoral fellowship of the Research Foundation—Flanders (FWO, GP.026.06N) and the Research Fund of Katholieke Universiteit Leuven (F/06/024).

Correspondence concerning this article should be addressed to Bouwien C. M. Smits-Engelsman, Faculty of Kinesiology and Rehabilitation Sciences, Katholieke Universiteit Leuven, Tervuursevest 101, 3001 Heverlee, Belgium. Electronic mail may be sent to bouwiensmits@hotmail.com.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7904-0004

Not only in visual tracking but also in manual tracking, the adequate formation of a predictive model of the target's motion is of the utmost importance, especially when the target is moving fast as the online use of feedback is impossible under these circumstances. To be successful, the future motion of the target has to be predicted and anticipated in the movements of the arm. Because manual tracking of a moving object is much more complex than visual tracking, the development of this ability takes more time.

Mounoud et al. (1985) examined the development of manual tracking in 5- to 9-year-old boys and concluded that 5-year-old children were able to produce a sinusoidal motion, but failed to create a global motor plan, including all characteristics of the target's motion. Often, a sequential pursuit strategy was observed at this age, in which short parts of the trajectory were planned and executed successively. The authors stated that the central representation of the target's motion became more and more elaborated with increasing age. They also found that low-speed targets were tracked successfully at a young age (above 5 years) but tracking high-speed targets improved until the age of 9 years. Zanone (1990) instructed 6- to 15-year-old children to manually track a visual target and found that the younger children relied mainly on the use of visual feedback to correct their movements, whereas the 15-year-olds adopted a feedforward strategy using a predictive model of the target's motion (see also Davidson & Wolpert, 2005).

To examine the ability of children and adolescents to quickly adjust the internal representation of the required movement of the pen tip, we designed a dynamic manual tracking task in which the target, which moved along a circular path, kept accelerating online when it was successfully tracked. The principal question was how the use of feedforward control developed throughout childhood and adolescence and whether tracking strategies changed with increasing age. For smooth *visual* pursuit of accelerating targets, it was recently shown that adults are able to use predictions of future target motion (Bennett & Barnes, 2006). Participants could scale their eye velocity appropriately to the expected target velocity after a short occlusion of the target but only when target velocity at reappearance was predictable (blocked presentations). To our knowledge, no studies have been performed in which the ability to *manually* track an accelerating target was examined in children and adolescents. Tracking an accelerating target requires the use of an ingenious predictive model of the target's motion, including not only the path along which the target moves and the velocity of the target but also its acceleration, to remain successful at higher target velocities (Magdaleno, Jex, & Johnson, 1970; Mounoud et al., 1983). It also requires a fast, continuous updating of this predictive model. Such a complex task is particularly useful to study the development of feedforward control.

It must be noted that at the start of a trial, the target moved at a low velocity, so that the task could be performed by using online visual feedback about position and velocity error to make corrections. However, with very fast moving targets, participants had to rely on feedforward control. It remained possible to perceive whether the cursor was inside or outside the target, but successful corrections based on that feedback information were not possible due to the visual–motor delay. The target was simply moving too fast. The correction would be made too late and the target would have decelerated as a result of bad tracking performance. The predictive model of the target's motion should include not only the future position of the target but also the future velocity and acceleration. On the basis of this model, the general motor plan was generated when the use of online feedback to generate corrections was impossible. Naturally, both visual and proprioceptive information about the movements that were made during task performance remained important sources regarding the target's trajectory, velocity, and acceleration. This information, gathered throughout the performance of the task, could be used to adjust the internal representation of the required movement of the pen tip to improve task performance in the rest of the trial and subsequent trials. The more appropriate this internal representation was, the better the task performance.

To test whether participants used online feedback during our tracking task, the number of submovements was determined by counting the number of velocity peaks per cm. In a study with monkeys (Roitman, Massaquoi, Takahashi, & Ebner, 2004), it was found that submovements during visually guided tracking reflected feedback-based corrective actions because they were highly correlated with specific error signals. Another strategy used to exploit feedback is the so-called step-and-hold strategy, in which movement velocity decreases to (close to) zero between submovements (Miall, Weir, & Stein, 1985, 1986; but see also Van der Meulen et al., 1990, for conflicting results). After a movement has been made to catch up with the target (a "step"), a stop ("hold") is made to determine what the next movement has to be. The participant is unable to pursue the target in a smooth way by predicting the movement of the target and anticipating to that. Mounoud et al. (1985) called it a sequential pursuit strategy, in which short parts of the trajectory are planned and executed successively. When the number of velocity peaks per cm and the number of stops are lower, a larger part of the tracking movement is made under feedforward control as fewer submovements are made. The movements are performed more smoothly.

We assumed that older children would be able to track the accelerating target at higher velocities than the younger ones because they will rely less on online feedback control and show a more optimal use of feedforward (or prospective) control based on a predictive model of the target's motion. This will result in higher maximum target velocities with increasing age, more circles drawn, less stops, and more fluent movements. We hypothesized that the rate of improvement would be highest at a younger age (Benguigui & Ripoll, 1998), leading to quadratic trends in the data (see also the Dependent Variables and Statistical Analysis section).

Method

Participants

Participants (N = 117, 62 girls, 55 boys, all Caucasian, mixed socioeconomic status [SES]) were recruited at two primary schools and a secondary school in Flanders, Belgium. First, the three schools that were willing to participate were visited to explain the study. Then, the parents of all children attending these schools were informed in writing about the study and they were asked for their permission. We aimed at including 10–15 boys and 10–15 girls of each of the five age groups we defined (6–7, 8–9, 10–11, 12–14, and 15–17 years). For each age group–gender combination, a random selection was made (if possible) from the group of children whose parents agreed with their participation. The ultimate number of participants in each subgroup was dependent on the number of informed consents from the parents and on the available testing time. All children had normal or corrected-to-normal vision. To measure their general motor skills (manual dexterity, ball skills, and equilibrium), all children were tested on the Movement Assessment Battery for Children (Henderson & Sugden, 1992; Smits-Engelsman, 1998). The percentile scores of the participants above 12 years of age were based on the norms of 11- to 12-year-olds. Of the 74 primary school children, 6 scored below the 15th percentile, and of the 43 secondary school children, 14 scored below the 15th percentile (11 of them scored especially poor at the equilibrium tasks). They were not excluded from participation. No definite motor control problems were identified as none of the children scored at or below the 5th percentile. Additional information about these age groups is provided in Table 1. This study was approved by the local Ethics Committee of Biomedical Research at the Katholieke Universiteit Leuven and performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki. The parents of all participants gave written informed consent. All children gave their assent.

Task and Procedure

The task, programmed in OASIS (De Jong, Hulstijn, Kosterman, & Smits-Engelsman, 1996) consisted of manually tracking a red dot (diameter 2.7 cm) presented on a 19-in. flat screen computer monitor that was vertically mounted in front of the participant. A digitizing tablet (WACOM UD-1218-RE), used to record the participant's response, was horizontally mounted on the table top, directly in front of the participant. The child moved the cursor, a small yellow dot (diameter 0.53 cm) that was also displayed on the monitor, by moving a wireless, electronic pen of normal appearance and weight and with an inkless pen point on a white sheet of paper (A3 format), positioned on the digitizer. The pen had to be held in the hand with which the child usually writes. The 2D position of the pen on the surface of the digitizer was

Table 1
Age Group Information

			Age		M-ABC percentile score	
Age group (years)	Boys (*N*)	Girls (*N*)	*M*	*SD*	*M*	*SD*
6–7	10	13	7 years 1 month	6 months	58	27
8–9	15	12	9 years 1 month	6 months	52	25
10–11	12	9	11 years 1 month	7 months	36	27
12–14	11	15	13 years 6 months	11 months	37	31
15–17	7	13	15 years 11 months	7 months	32	30

Note. M-ABC = Movement Assessment Battery for Children.

sampled at a frequency of 206 Hz and with an accuracy of 0.1 mm. At the start of each trial, the target velocity was 20°/s. Target velocity never dropped below this value throughout the task. When tracked successfully (i.e., the cursor being completely contained within the target for at least 75% of a 0.5 s time slice period), the target slowly accelerated with an angular velocity increase of 5%. Otherwise, it quickly decelerated (i.e., the angular velocity decreased with 23%) to allow reentry. Although acceleration and deceleration occurred in discrete steps, these steps were small, such that the target's motion was not perceived as being jerky. Participants were instructed to try to keep the yellow dot on the red target. To make the task more appealing, the target represented a villain and the cursor a policeman who had to keep the villain captured by remaining "on top of" him. Performance feedback (i.e., the number of circles drawn) was given after each trial. Two conditions were presented in which the red dot rotated along a circular path with a diameter of either 3 or 10 cm, clockwise for right-handed participants and counterclockwise for left-handed participants. It must be noted that the change in linear velocity across time was larger in the large circle condition than in the small circle condition, as it is the angular velocity that changed. A target moving at an angular motion of 20°/s covered a larger linear distance in a large circle as compared to a small one. Furthermore, the above-mentioned diameters of the target, cursor, and the circular trajectory reflected the sizes on the digitizing tablet. On the computer monitor, the diameters were all approximately twice as large. After one practice trial (duration 60 s) with large circular movements, two trials lasting 60 s each were performed within each condition. Half of the participants started with large and the other half with small circular movements. Data collection started when the distance between the centers of cursor and the target was less than 4.0 cm for the first time. Participants were allowed to rest between trials. The experiment lasted typically 10 min.

Dependent Variables and Statistical Analysis

The data were analyzed using Matlab and the OASIS software (De Jong et al., 1996). The 2D positional data of the pen were low-pass filtered using a zero-phase lag, second-order Butterworth filter with a cutoff frequency of 10 Hz and were subsequently differentiated to calculate movement velocity and acceleration (Matlab). The dependent variables were subdivided into measures that reflected tracking success and the extent to which feedback was used during tracking.

Measures of tracking success consisted of the number of circles completed during a trial and the maximum velocity of the target. These measures reflected to what extent a participant succeeded in unremittingly keeping the cursor within the target, even at higher target velocities. When tracking success was great, the target kept accelerating, resulting in more circles drawn and a higher target velocity. Additional measures were the median of the distance between the center of the target and the center of the cursor throughout a trial, the total amount of time that the cursor was on top of the target, and the median of the gain (i.e., the velocity of the cursor divided by the velocity of the target throughout a trial). These three measures also reflected how well a participant tracked the target. However, the velocity the target reached, which was an important indicator of the degree to which feedforward control was used, could not be inferred from these measures. To clarify this, note the following example. If someone could perfectly track the target, but only at lower velocities, the cursor would have left the target sometimes, namely whenever the velocity became too high, resulting in a decelerating target. The period that the cursor was outside of the target might be very short. The median of the distance between the centers could still be low under these circumstances, the duration within the target long, and the gain close to one.

To find out to which degree online feedback was used, we determined (a) the number of peaks per cm (i.e., normalized for the distance covered) in the velocity profile of the pen trace (in which a peak is defined as a point where the slope of the velocity curve changed sign), which corresponds to the number of submovements assumed to indicate corrective actions and (b) the number of times that the velocity of the pen was below 0.5 cm/s, reflecting the extent to which a feedback-based step-and-hold strategy was used.

Statistical analysis consisted of regression analyses and repeated measures analyses of variance (ANOVAs). Regression analyses were performed to reveal possible linear and quadratic trends across age groups. A linear relationship indicates a linear improvement of task performance with age. However, when the quadratic trend is also significant, this can point toward a later start of the improvement, a "leveling off" or stop of this improvement at a certain age, or even at a deterioration of task performance after an initial improvement or an improvement after an initial deterioration. We expected both linear and quadratic trends for our dependent variables. Children will evidently improve with age, but the improvement will be largest at a younger age (see also Benguigui &

Ripoll, 1998). Repeated measures ANOVAs were performed to test the effects of the within-subjects factors movement condition (two levels: small and large circular movements) and trial number (two levels: first and second trial within a condition), the effect of the between-subjects factor age group (five levels: 6–7, 8–9, 10–11, 12–14, and 15–17 years), and the interactions between these factors. So-called repeated contrasts were performed to test differences between successive age groups. An alpha level of .05 was used for all statistical tests.

Results

In Figures 1 and 2, some examples of task performance are displayed for a 7-year-old and a 16-year-old boy. For both conditions (big and small circles), the older boy drew more circles than the younger boy, as can be seen in the top two panels of each figure (X and Y location), which implies a better tracking performance by the older boy. The distance between the centers of target and cursor remained below 1 cm during the larger part of the trial (third panels). Finally, in the lower panels, it can be seen that the maximum target velocity (black line) was higher, whereas the number of stops (i.e., pen velocity below 0.5 cm/s) was lower for the older boy.

The means and standard deviations of the various dependent variables across the five age groups are presented in Table 2.

Number of Circles Drawn

Both a linear trend, $R^2 = .47$, $F(1, 115) = 102.91$, $p < .001$, and a quadratic trend, $R^2 = .51$, $F(2, 114) = 58.59$, $p < .001$, were found across age groups for the number of circles drawn. As can be seen in Figure 3A, the number of circles increases up to a certain age after which it levels off. This finding was confirmed by the ANOVA, which showed a main effect of age, $F(4, 112) = 29.43$, $p < .001$. The subsequent repeated contrasts indicated that only the differences between the 6- to 7-year-olds and the 8- to 9-year-olds and between the 8- to 9-year-olds and the 10- to 11-year-olds were significant ($ps < .001$).

To determine the effects of circle diameter and trial number, the data were first corrected for the difference in circumference of the drawn circles (circumference = diameter × π; i.e., for the large and small circles, it was 10π and 3π, respectively), by multiplying the amount of drawn circles in the big circle condition by 3.33. We found a main effect of diameter, $F(1, 112) = 971.67$, $p < .001$ (small circle condition: 25.1 circles drawn; big circle condition: 44.9 circles drawn), but there was also a significant interaction between diameter and age group, $F(4, 112) = 13.70$, $p < .001$. When the increase in the number of circles drawn between the conditions with small and large circular movements was expressed as a percentage (an increase of 100% means a doubling), the mean increase changed with age. For Age Groups 1–5, the increases were 63.1%, 76.6%, 78.7%, 79.9%, and 85.3%, respectively. The repeated contrasts between successive age groups showed similar age effects as compared to the overall analyses (Table 2) in the large circles and the small circles condition. There was also an effect of trial number, $F(1, 112) = 6.09$, $p = .015$, and a Diameter × Trial Number interaction, $F(1, 112) = 4.23$, $p = .042$, indicating that there was only an increase in the number of circles drawn between the first and the second trial (from 44.2 to 45.7 circles) in the big circle condition.

Maximum Target Velocity

A linear trend, $R^2 = .29$, $F(1, 115) = 47.79$, $p < .001$, as well as a quadratic trend, $R^2 = .33$, $F(2, 114) = 27.56$, $p < .001$, were found for maximum target velocity. Maximum target velocity increased with age from 8.9 until 13.4 cm/s, $F(4, 112) = 13.94$, $p < .001$, but contrasts indicated that only the performance difference between the 6- to 7-year-olds and the 8- to 9-year-olds and between the 8- to 9-year-olds and the 10- to 11-year-olds was significant ($ps = .015$ and .003, respectively).

Participants reached higher maximum target velocities when large circular movements had to be drawn (14.2 cm/s), instead of small circular movements (9.2 cm/s), $F(1, 112) = 326.55$, $p < .001$. There was also an interaction between diameter and age group, $F(4, 112) = 2.67$, $p = .036$. Repeated contrasts between successive age groups, for the small and the large circular movements separately, resulted in p values similar to the overall analyses (see Table 2), with the exception of the difference between Age Groups 1 and 2 in the small movement condition. This difference just failed to reach significance ($p = .054$).

Median of the Distance Between Centers of Cursor and Target

Again, both a linear trend, $R^2 = .12$, $F(1, 115) = 15.65$, $p = .001$, and a quadratic trend, $R^2 = .23$, $F(2, 114) = 17.05$, $p < .001$, were found (Figure 3B). The results of the ANOVA showed a main effect of age, $F(4, 112) = 8.76$, $p < .001$. The median of the distance

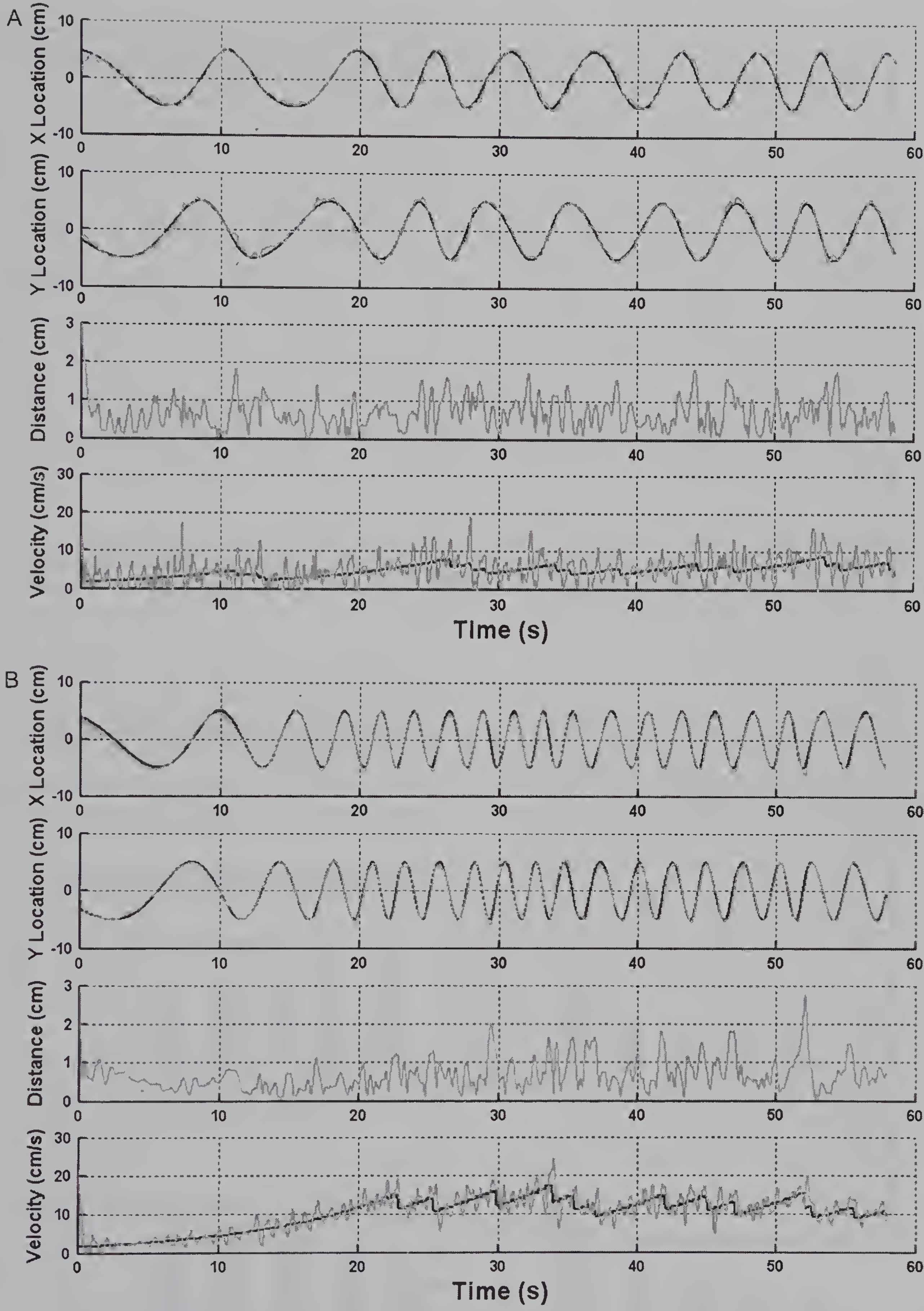


Figure 1. Typical examples of task performance in the *big* circles condition: (A) 7-year-old boy and (B) 16-year-old boy.
Note. In the top two panels of Figures 1A and 1B, the horizontal and vertical position of the centers of the target (black) and cursor (gray) throughout task performance are depicted. In the third panel, the distance between the centers of the target and the cursor is displayed, whereas in the last panel, the velocities of the target (black) and cursor (gray) are shown.

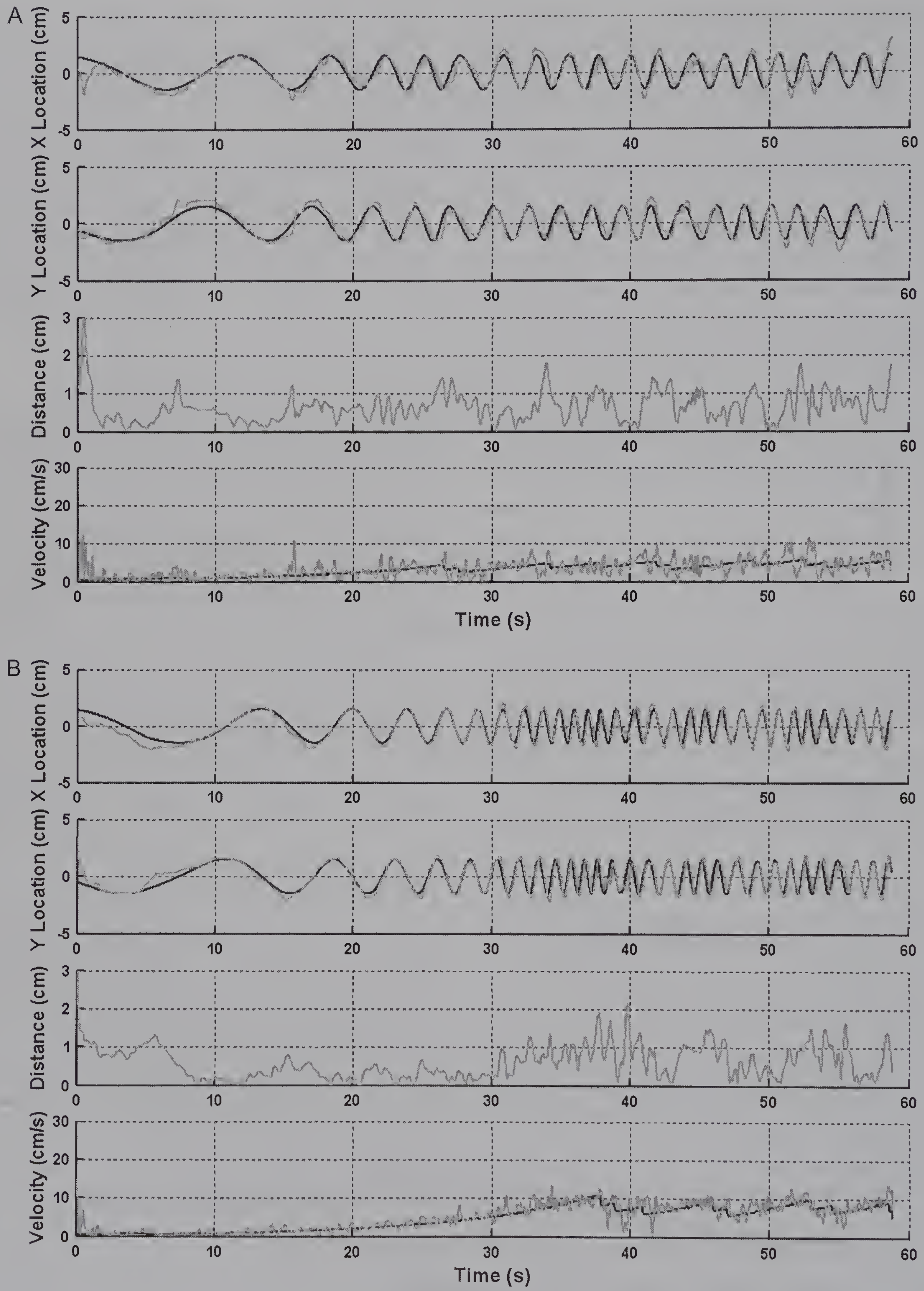


Figure 2. Typical examples of task performance in the *small* circles condition: (A) 7-year-old boy and (B) 16-year-old boy.
Note. In the top two panels of Figures 2A and 2B, the horizontal and vertical position of the centers of the target (black) and cursor (gray) throughout task performance are depicted. In the third panel, the distance between the centers of the target and the cursor is displayed, whereas in the last panel, the velocities of the target (black) and cursor (gray) are shown.

Table 2

Means and Between-Subjects Standard Deviations of the Dependent Variables

	6–7			8–9			10–11			12–14			15–17		
Age group	*M*	*SD*		*M*	*SD*		*M*	*SD*		*M*	*SD*		*M*	*SD*	Group effect
No. of circles	13.8	3.0	***	17.2	3.6	***	21.0	2.4		21.9	3.8		22.9	3.5	$F(4, 112) = 29.43, p < .001$
Maximum velocity target (cm/s)	8.9	2.5	*	10.7	2.9	**	12.8	1.8		13.1	2.5		13.4	2.4	$F(4, 112) = 13.94, p < .001$
Median distance between centers (cm)	0.61	0.07	***	0.57	0.03	*	0.54	0.03	*	0.56	0.04		0.56	0.03	$F(4, 112) = 8.76, p < .001$
Total duration in target (s)	48.4	2.1	***	50.0	0.6	†	50.6	0.9		50.6	0.7		50.6	0.7	$F(4, 112) = 17.07, p < .001$
Median gain	1.000	0.05		0.997	0.07		0.984	0.05		0.998	0.07		1.012	0.05	*ns*
No. of velocity peaks/cm	0.7	0.3		0.7	0.2	*	0.6	0.1		0.5	0.1	†	0.4	0.1	$F(4, 112) = 9.42, p < .001$
No. of stops	41.0	17.7	***	26.8	9.2	*	20.0	5.2		15.7	4.4		12.4	3.6	$F(4, 112) = 31.47, p < .001$

$^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

between the centers of target and cursor within a trial initially decreased with age but subsequently increased; it decreased between ages 6–7 and 8–9 ($p < .001$) and between ages 8–9 and 10–11 ($p = .019$), but it increased again between ages 10–11 and 12–14 ($p = .049$). No difference was found between the two oldest age groups. A main effect of diameter was found, $F(1, 112) = 92.89$, $p < .001$, showing that the median of the distance between the centers was larger for the large circular movements (0.61 cm) than for the small circular movements (0.55 cm). Besides the main effect of trial number, $F(1, 112) = 7.05$, $p = .009$, the Trial Number × Age Group interaction was also significant, $F(4, 112) = 3.19$, $p = .016$. The distance decreased with practice (from the first to the second trial), but only in age group 1 (from 0.62 cm to 0.60 cm; $p = .007$) and 3 (from 0.55 cm to 0.53 cm; $p = .001$).

Total Duration in Target

Regression analyses revealed both a linear trend, $R^2 = .26$, $F(1, 115) = 40.55$, $p < .001$, and a quadratic trend, $R^2 = .37$, $F(2, 114) = 33.02$, $p < .001$, showing that the amount of time the cursor was inside the

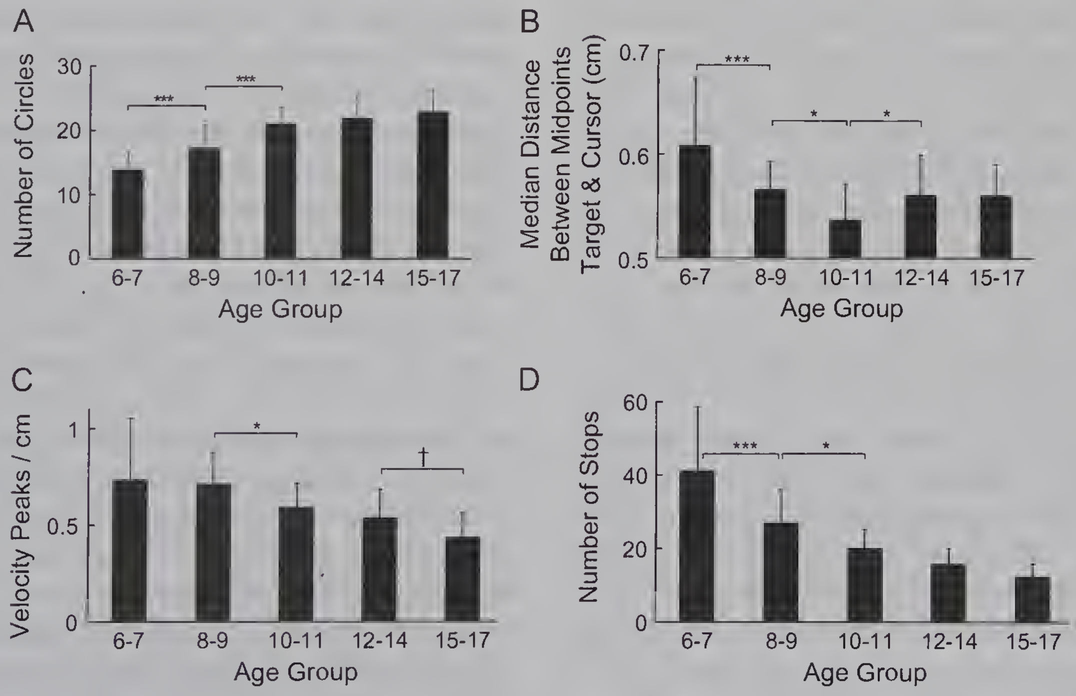


Figure 3. Effects of age group on (A) the number of circles drawn, (B) the distance between the center of the target and the center of the cursor in cm, (C) the number of velocity peaks per cm in the velocity profile of the pen, and (D) the number of stops (velocity < 0.5 cm/s).
Note. Error bars reflect the between-subjects standard deviations.
$\dagger p < .10$. $^{*}p < .05$. $^{***}p < .001$.

target decreased in the youngest age groups, whereas it remained stable in the older age groups. Contrast analyses of the main effect of age, $F(4, 112) = 17.07$, $p < .001$, revealed that only the difference between the first two age groups was significant ($p < .001$), whereas the difference between ages 8–9 and 10–11 only reached marginal significance ($p = .062$). The cursor was in the target for a longer duration when the circles were small (51.2 s) as compared to large (48.9 s), $F(1, 112) = 298.86$, $p < .001$. Furthermore, the duration in the target was slightly longer in the second trial as compared to the first trial (49.8 vs. 50.2 s), $F(1, 112) = 12.59$, $p = .001$.

Median Gain

No effect of age group was found, neither with regression analyses nor with ANOVA. Averaged across all participants, the median gain was 0.998 ($SD = 0.061$).

Number of Velocity Peaks per Centimeter

As shown by the linear trend, $R^2 = .24$, $F(1, 115) = 36.89$, $p < .001$, and the quadratic trend, $R^2 = .25$, $F(2, 114) = 18.57$, $p < .001$, the number of velocity peaks per cm started to decrease at the age of 8–9 (Figure 3C). Contrast analyses of the main effect of age, $F(4, 112) = 9.42$, $p < .001$, showed a significant decrease in the amount of submovements in children ages 10–11 as compared to children ages 8–9 years ($p = .033$), whereas the decrease between ages 12–14 and 15–17 almost reached significance ($p = .071$). Movements were performed more smoothly in the large circle condition (0.34 velocity peaks per cm) than in the small circle condition (0.87 velocity peaks per cm), $F(1, 112) = 422.05$, $p < .001$.

Number of Stops

Both a linear trend, $R^2 = .48$, $F(1, 115) = 106.47$, $p < .001$, and a quadratic trend, $R^2 = .52$, $F(2, 114) = 62.79$, $p < .001$, were found across age groups, which indicates that the number of stops decreased with age but also that this decrease leveled off in the older age groups (Figure 3D). Contrast analyses of the main effect of age, $F(4, 112) = 31.47$, $p < .001$, revealed again that only the differences between the 6- to 7-year-olds and 8- to 9-year-olds and between the 8- to 9-year-olds and the 10- to 11-year-olds were significant ($p < .001$ and $p = .016$, respectively).

When the circular movements were large, participants stopped on average 12.7 times during task performance, whereas they stopped 34.2 times when small circular movements had to be drawn, $F(1, 112) = 652.51$, $p < .001$. The interaction between diameter of the circles and age group was, however, also significant, $F(4, 112) = 6.44$, $p < .001$; the effect of circle diameter was significant in all age groups, but the relative difference in the number of stops between large and small circle conditions (i.e., the increase expressed as a percentage of the number of stops in the large circle condition) increased with age. For Age Groups 1–5, the increases were 85.8%, 181.1%, 289.5%, 285.2%, and 267.5%, respectively). Age effects appeared to be similar in the large and the small circle condition. The repeated contrasts between successive age groups, for the small and the large circular movements separately, showed age effects similar to the overall analyses (Table 2), although an additional trend was found: In the small circles condition, the difference between the 10- to 11-year-olds and the 12- to 14-year-olds just failed to reach significance ($p = .052$).

Discussion

The aim of this study was to determine the development of feedforward control in children and adolescents during a dynamic manual tracking task and the ability to quickly adjust the predictive model of the target's motion. We examined the strategies children use when they have to flexibly adapt their movements to the changing velocities of the target while preserving accuracy (keeping the cursor within the target area). The number of circles and the maximum target velocity increased with age, whereas the number of velocity peaks and the number of stops decreased with age. We therefore conclude that a feedback-based strategy was gradually replaced by a more effective feedforward-based strategy to remain able to track high-speed targets.

All children were able to track the target, but the younger children (6- to 9-year-olds) could track the target only at lower velocities, whereas the older children and adolescents still tracked the target successfully at higher velocities, resulting in more completed circles. As can be concluded from the number of stops the participants made during tracking, the younger children made more use of a step-and-hold strategy (feedback based), which is only appropriate at lower velocities. With increasing age, the use of this strategy decreased (i.e., the number of stops decreased), whereas the use of a predictive model of the target's motion increased (e.g., higher maximum target velocities and less velocity peaks/cm). The linear decrease in the number of stops and the number of velocity peaks per cm continued until 15 – 17 years (see Table 2), although no longer significantly. Also note the finding that the number of velocity peaks per cm was similar in children ages 6 – 7 and 8 – 9 years but that the number of stops was significantly higher in the first age group. These developmental trends are hypothesized to be consistent with the unfolding of feedforward control processes.

Although speculative, we suggest three phases in the development of tracking control based on our results: (a) learning to stay on top of the target, (b) anticipating the movement of the target so that the target can be tracked at higher velocities, and (c) learning to rely less on online visual feedback, resulting in fewer submovements and increasingly higher possible target velocities. Our youngest participants ages 6 – 7 years were already able to stay on top of the target during the larger part of a 1-min trial. Moreover, the median distance between the centers of target and cursor and the median gain did not show large differences with those of the older age groups. However, the maximum target velocity increased up to ages 10 – 11, and decreases in the amount of submovements were found up to ages 15 – 17.

Various factors might have played a role in causing the age-related differences in task performance. In our view, the most important factor is the gradual improvement of feedforward processes over the course of development. The finding that the velocity of the target increased with age, up to a value at which the use of feedback was severely constrained—given the visual – motor delay—suggests that the part of the tracking movement that is made under feedforward control increased with age (see also the next paragraph). The younger children showed increased tracking difficulties at high target velocities, probably because they were unable to switch to feedforward control. The finding that the number of submovements decreased with age, as indicated by a decreasing number of velocity peaks per centimeter and stops, supports this. The movements became smoother with development. Feedforward processes can be subdivided in several separate processes that could have played a role during the performance of our tracking task: first, the use of predictive models of external object dynamics (i.e., the target's motion) and second, the use of predictive models of the dynamics of the arm (forward and inverse internal models; Davidson & Wolpert, 2005; Wolpert, Ghahramani, & Flanagan, 2001; Wolpert, Miall, & Kawato, 1998; for kinematic data that suggest the use of these forward and inverse internal models by children and the improvement of their quality with development, see Contreras-Vidal, 2006; Jansen-Osmann, Richter, Konczak, & Kalveram, 2002). Forward internal models predict the sensorimotor consequences of actions and the future state of the moving limb by using copies of motor commands (efference copies). This permits fast adjustments to the unfolding movement and reduces dependence on slower feedback control. Inverse internal models provide the motor command necessary to achieve some desired trajectory. Studies into the neural correlates of tracking behavior suggest that the cerebellum provides the predictive internal models necessary for the coordination between eye and hand movements during tracking (e.g., Miall & Jenkinson, 2005; Miall, Reckess, & Imamizu, 2001; Roitman, Pasalar, Johnson, & Ebner, 2005; Wolpert et al., 1998).

A second factor that might have played a role is the increased speed of information processing in older children enabling them to use feedback about position and velocity error at higher target velocities than younger children (Bourgeois & Hay, 2003; Wallace, Newell, & Wade, 1978; Wickens, 1974). Keele and Posner (1968) concluded that it takes about 200 ms to use visual feedback to correct aiming movements. Although Carlton (1981) found that the time between the presentation of visual error information and the initiation of a movement correction was only 135 ms, it must be noted that in our task, the target moved and its velocity changed online; accordingly, it would probably take more time to process visual feedback. Although it is merely speculation because it was not our aim to determine the visual – motor delay, we assume that this delay is at least 200 ms when performing our task.

We can approximately determine the maximum target velocity for the participant to be able to successfully use visual feedback to correct the tracking error. Imagine the situation in which the cursor is just inside the target and does not move while the target moves across this cursor. Then, the distance that the target travels between the first and the last moment that the cursor is completely inside the target area is 2.70 cm (the diameter of the target)—0.53 cm (the diameter of the cursor) = 2.17 cm (when we assume, to simplify matters, that the target moves on a straight line instead of in circles). The time interval during which the target covers this distance should be no longer than the visual – motor delay. When 2.17 cm is divided by the assumed visual – motor delay of 200 ms, it can be concluded that the maximum target velocity for the use of visual feedback to be possible was 10.9 cm/s, which approximately corresponded to the average maximum velocity the target reached in the age group of 8- to 9-year olds.

Hence, 6- to 9-year-old children might have used visual feedback to minimize the difference between the target and the cursor position. The average maximum target velocity in the 10- to 17-year-old participants varied between 12.8 and 13.4 cm/s. If they use only visual feedback to correct errors, the visual–motor delay should be around 165 ms, which might be too short a period given the complexity of our task. This might provide another indication that the older children use feedforward control when the target reaches velocities above 11 cm/s.

A third factor that might explain the age-related differences in task performance is the decreasing variability in motor output with increasing age (e.g., Schneiberg, Sveistrup, McFadyen, McKinley, & Levin, 2002) which simplifies the adequate prediction of the future position of the moving limb. As a result of this improved muscle coordination, reflected in more fluent movements, older children are also better able to make fast adjustments in movement velocity.

A fourth factor is the improvement of visual pursuit with age. In studies of oculomotor functioning, clear developmental trends were found, but the age at which visual tracking reaches adult levels is higher as compared to manual tracking. Ross, Radant, and Hommer (1993) found improvements in visual tracking of a target until the age of 15, but no older participants were tested. Katsanis, Iacono, and Harris (1998) found that during preadolescence (11–12 years of age), the oculomotor system is not working optimally, but by late adolescence (17–18 years of age), smooth pursuit eye tracking has reached adult levels of performance. When the visual tracking of a target improves, it is likely that manual tracking also improves because these movements are tightly coupled in a manual tracking task.

A final factor of importance is the ability to perceive acceleration. Unfortunately, to our knowledge, no studies have been done into developmental changes in children with regard to visual perception of acceleration. In a study with adults, Rosenbaum (1975) showed that acceleration can be perceived accurately, although this ability deteriorated with increasing acceleration. Also, results of Calderone and Kaiser (1989) suggest that adults are relatively sensitive to linear accelerations. Schmerler (1976) found that deceleration was easier to perceive than acceleration, but Calderone and Kaiser reported that this was dependent on the direction of motion. In addition, Schmerler, but also Gottsdanker, Frick, and Lockard (1961), reported that a gradual change in velocity (as in our study) is more difficult to detect than an abrupt change. If the ability to perceive acceleration is not yet developed in young children, then this might be a factor that influences the age-related differences in tracking performance that we found. However, it must be noted that besides the visual information about target acceleration, our participants also had proprioceptive information. They could have felt the acceleration of their moving hand while performing the tracking task. Von Hofsten and Rösblad (1988) showed that children between 5 and 12 years of age greatly improve on a pointing task in which they had to rely only on proprioceptive information.

We have considered the possibility that differences in fine motor control of the hands are a larger part of the explanation for the age-related differences that we found than differences in control strategies, especially for the small circle condition, as finger movements are more important in this condition than in the large circle condition. We examined the Age Group × Circle Size interactions that we found. If differences in fine motor control constituted a large part of the explanation for the age-related differences in our data, one would expect a larger difference between the youngest and the oldest children in the small circle condition than in the large circle condition. This did not appear to be the case (Table 3). The relative differences between the age groups were larger in the big circle condition, for the performance of which fine motor control of the hands was less necessary.

Table 3
Absolute and Relative Differences Between the Youngest and the Oldest Age Group for the Big and Small Circles Conditions Separately

		6–7 years	15–17 years	Absolute difference	Relative difference (%)
No. of circles	Big (corrected)	30.2	54.5	+24.3	+80
	Small	18.7	29.5	+10.8	+58
Maximum target velocity (cm/s)	Big	10.7	16.5	+5.8	+54
	Small	7.2	10.3	+3.1	+43
No. of stops	Big	28.7	5.3	−23.4	−82
	Small	53.4	19.5	−33.9	−63

Note. Differences were only calculated for the dependent variables for which age group × circle size interactions were found.

When we take stock of the complete set of results, we conclude that differences in control strategies are a more likely explanation than the differences in fine motor control of the hands.

Some other interesting results were found with respect to small versus large circular movements. Apparently, it was easier to track the target when it made the large as compared to the small circular movements. More circles were drawn and a higher maximum target velocity was reached. This might have been caused by the fact that the small circle has a much larger curvature in a smaller area. When we would have scaled the size of the target and, hence, the margin of error, for this difference in difficulty, the effects of circle diameter might have been absent or at least smaller, but the developmental trends would not have been different because all participants performed the task in the same two conditions. Furthermore, a difference in the amount of time available for the use of online feedback would have been introduced under these circumstances.

The tracking movement was also performed more smoothly (i.e., fewer submovements per centimeter) and fewer stops were made in the large circle condition. A possible explanation for these findings might be the difference in musculature used. When drawing large circles, the proximal joints (i.e., the elbow and shoulder) are primarily activated, whereas the distal joints (i.e., fingers and wrist) become more prominent when drawing smaller circles. The smoothness of the movement will benefit from the greater mass of the elbow and shoulder that will have a damping effect on the movement. In addition, the stronger muscles surrounding the more proximal joints generate less variable torques (Hamilton, Jones, & Wolpert, 2004). At the distal joints, the noise in the motor output is greater because they are acted upon by weaker muscles, leading to more variable movements. As a result of the higher inertia of the more proximal body segments, it is also more difficult to come to an almost complete stop during task performance when fast elbow and shoulder flexion–extension movements are performed as compared to flexion–extension movements of the fingers. These factors can also explain that the younger children show a relatively smaller increase in number of stops between the large and the small circle condition. Younger children have shorter fingers, so that they need to use their elbow and shoulder more when drawing small circles, whereas the older children have the possibility to use mainly their fingers and wrist to move the pencil.

Feedforward or prospective control is extremely important in daily life. Von Hofsten (1993) argued that "the development of action is basically a matter of acquiring prospective control" (p. 254) and that "actions constitute dynamic interactions between an organism and the outside world which, by necessity, have to be future-oriented" (p. 268). We have to make predictions about what is going to happen next in the outside world and adjust our movements accordingly. Although very young infants appear to be able to visually track a moving object and make inferences about the motion of objects that move behind occluders (Spelke et al., 1994), developmental changes can be observed into childhood when the task becomes more complex, as in our study. Participants not only had to visually track the target but also manually, and in addition, the target accelerated, resulting in a very complex predictive model of the target's motion required for successful tracking.

We assume that performance on our dynamic manual tracking task corresponds to performance of daily life motor activities in which prediction of object motion and anticipation of the person's movements to that object's motion is necessary. An example is catching or hitting fast approaching balls. Lefebvre and Reid (1998) showed that young children (ages 5–6 years) are worse in predicting ball flight as compared to older children (ages 7–12 years) at short viewing times. In a study of Benguigui and Ripoll (1998), tennis players and novices, 7, 10, 13, and 23 years of age, were asked to perform a task in which they had to predict the moment at which a moving object (with a constant velocity or a constant acceleration or a constant deceleration) would arrive on a target. Main improvements of timing accuracy were found between the ages of 7 and 10 years, which is comparable to the present findings. In addition, playing tennis ameliorates timing accuracy; the development of timing accuracy is accelerated.

To determine the quality of the internal representation of the required movement of the pen tip more directly, future studies might include a condition in which the target's visibility is intermittent, thereby forcing participants to rely on the predictive model of the target's motion. Velocity and acceleration of the target should not be dependent on tracking performance under these conditions. Otherwise task performance would become impossible. It might be an idea to first examine tracking performance with a continuously visible accelerating target so that participants can build the predictive model. In a subsequent trial, the target can disappear after some time so that reliance on the predictive model is required.

To examine the possibility that young children would have learned to perform the task after more practice, an experiment is required in which children perform the task many times. It might be that it just takes more time to build an adequate predictive model of the target's motion in younger children but that they can do it if allowed enough practice.

To summarize, we have provided evidence for the gradual maturation of feedforward control processes with increasing age. While growing up, children learn to exploit the advantages of feedforward control to be able to flexibly adapt their movements to the task demands. They gain experience in building a predictive model of both the target's motion and the dynamics of the arm, resulting in more efficient strategies.

References

Benguigui, N., & Ripoll, H. (1998). Effects of tennis practice on the coincidence timing accuracy of adults and children. *Research Quarterly for Exercise and Sport, 69*, 217–223.

Bennett, S. J., & Barnes, G. R. (2006). Smooth ocular pursuit during the transient disappearance of an accelerating visual target: The role of reflexive and voluntary control. *Experimental Brain Research, 175*, 1–10.

Bertenthal, B. I., Longo, M. R., & Kenny, S. (2007). Phenomenal permanence and development of predictive tracking in infancy. *Child Development, 78*, 350–363.

Bourgeois, F., & Hay, L. (2003). Information processing and movement optimization during development: Kinematics of cyclical pointing in 5- to 11-year-old children. *Journal of Motor Behavior, 35*, 183–195.

Calderone, J. B., & Kaiser, M. K. (1989). Visual acceleration detection: Effect of sign and motion orientation. *Perception and Psychophysics, 45*, 391–394.

Carlton, L. G. (1981). Processing visual feedback information for movement control. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 1019–1030.

Contreras-Vidal, J. (2006). Development of forward models for hand localization and movement control in 6- to 10-year-old children. *Human Movement Science, 25*, 634–645.

Davidson, P. R., & Wolpert, D. M. (2005). Widespread access to predictive models in the motor system: A short review. *Journal of Neural Engineering, 2*, S313–S319.

Davol, S. H., Hastings, M. L., & Klein, D. A. (1965). Effect of age, sex, and speed of rotation on rotary pursuit performance by young children. *Perceptual and Motor Skills, 21*, 351–357.

De Jong, W. P., Hulstijn, W., Kosterman, B. J. M., & Smits-Engelsman, B. C. M. (1996). OASIS software and its applications in experimental handwriting research. In M.L. Simmer, C. G. Leedham, & A. J. W. M. Thomassen (Eds.), *Handwriting and drawing research: Basic and applied issues* (pp. 429–440). Amsterdam: IOS.

Dunham, P., Allan, R., & Winter, R. (1985). Tracking ability of elementary school-age children. *Perceptual and Motor Skills, 60*, 771–774.

Gottsdanker, R., Frick, J. W., & Lockard, R. B. (1961). Identifying the acceleration of visual targets. *British Journal of Psychology, 52*, 31–42.

Gredebäck, G., & von Hofsten, C. (2004). Infants' evolving representations of object motion during occlusion: A longitudinal study of 6- to 12-month-old infants. *Infancy, 6*, 165–184.

Gredebäck, G., von Hofsten, C., Karlsson, J., & Aus, K. (2005). The development of two-dimensional tracking: A longitudinal study of circular pursuit. *Experimental Brain Research, 163*, 204–213.

Hamilton, A. F. de C., Jones, K. E., & Wolpert, D. M. (2004). The scaling of motor noise with muscle strength and motor unit number in humans. *Experimental Brain Research, 157*, 417–430.

Henderson, S. E., & Sugden, D. A. (1992). *Movement assessment battery for children: Manual*. London: Psychological Corporation.

Jansen-Osmann, P., Richter, S., Konczak, J., & Kalveram, K.-T. (2002). Force adaptation transfers to untrained workspace regions in children: Evidence for developing inverse dynamic motor models. *Experimental Brain Research, 143*, 212–220.

Katsanis, J., Iacono, W. G., & Harris, M. (1998). Development of oculomotor functioning in preadolescence, adolescence, and adulthood. *Psychophysiology, 35*, 64–72.

Keele, S. W., & Posner, M. I. (1968). Processing visual feedback in rapid movement. *Journal of Experimental Psychology, 77*, 155–158.

Lefebvre, C., & Reid, G. (1998). Prediction in ball catching by children with and without a developmental coordination disorder. *Adapted Physical Activity Quarterly, 15*, 299–315.

Magdaleno, R. E., Jex, H. R., & Johnson, W. A. (1970). Tracking quasi-predictable displays: Subjective predictability gradations, pilot model for periodic and narrow band inputs. In *Proceedings of the 5th Annual NASA-University Conference on Manual Control* (NASA-SP215, pp. 391–428). Washington, DC: NASA.

Miall, R. C., & Jenkinson, E. W. (2005). Functional imaging of changes in cerebellar activity related to learning during a novel eye-hand tracking task. *Experimental Brain Research, 166*, 170–183.

Miall, R. C., Reckess, G. Z., & Imamizu, H. (2001). The cerebellum coordinates eye and hand tracking movements. *Nature Neuroscience, 4*, 638–644.

Miall, R. C., Weir, D. J., & Stein, J. F. (1985). Visuomotor tracking with delayed visual feedback. *Neuroscience, 16*, 511–520.

Miall, R. C., Weir. D. J., & Stein, J. F. (1986). Manual tracking of visual targets by trained monkeys. *Behavioural Brain Research, 20*, 185–201.

Mounoud, P., Hauert, C. A., Mayer, E., Gachoud, J. P., Guyon, J., & Gottret, G. (1983). Visuo-manual tracking strategies in the three to five year-old child. *Archives de Psychologie, 51*, 23–33.

Mounoud, P., Viviani, P., Hauert, C. A., & Guyon, J. (1985). Development in visuomanual tracking in 5- to 9-year-old boys. *Journal of Experimental Child Psychology, 40*, 115–132.

Roitman, A. V., Massaquoi, S. G., Takahashi, K., & Ebner, T. J. (2004). Kinematic analysis of manual tracking in monkeys: Characterization of movement intermittencies during a circular tracking task. *Journal of Neurophysiology, 91*, 901–911.

Roitman, A. V., Pasalar, S., Johnson, M. T. V., & Ebner, T. J. (2005). Position, direction of movement, and speed tuning of cerebellar Purkinje cells during circular manual tracking in monkey. *Journal of Neuroscience, 25*, 9244–9257.

Rosenbaum, D. A. (1975). Perception and extrapolation of velocity and acceleration. *Journal of Experimental Psychology: Human Perception and Performance, 1*, 395–403.

Ross, R. G., Radant, A. D., & Hommer, D. W. (1993). A developmental study of smooth-pursuit eye-movements in normal children from 7 to 15 years of age. *Journal of the American Academy of Child and Adolescent Psychiatry, 32*, 783–791.

Schmerler, J. (1976). The visual perception of accelerated motion. *Perception, 5*, 167–185.

Schneiberg, S., Sveistrup, H., McFadyen, B., McKinley, P., & Levin, M. F. (2002). The development of coordination for reach-to-grasp movements in children. *Experimental Brain Research, 146*, 142–154.

Smits-Engelsman, B. C. M. (1998). *Movement assessment battery for children (handleiding)* [Dutch Manual]. Lisse, The Netherlands: Swets & Zeitlinger.

Spelke, E. S., Katz, G., Purcell, S. E., Ehrlich, S. M., & Breinlinger, K. (1994). Early knowledge of object motion: Continuity and inertia. *Cognition, 51*, 131–176.

van der Meulen, J. H. P., Gooskens, R. H. J. M., Willemse, J., Denier van der Gon, J. J., & Gielen, C. C. A. M. (1990). Arm-tracking performance with and without visual feedback in children and adults: Developmental changes. *Journal of Motor Behavior, 22*, 386–405.

von Hofsten, C. (1983). Catching skills in infancy. *Journal of Experimental Psychology: Human Perception and Performance, 9*, 75–85.

von Hofsten, C. (1993). Prospective control: A basic aspect of action development. *Human Development, 36*, 253–270.

von Hofsten, C. (2007). Action in development. *Developmental Science, 10*, 54–60.

von Hofsten, C., & Rosander, K. (1997). Development of smooth pursuit tracking in young infants. *Vision Research, 37*, 1799–1810.

von Hofsten, C., & Rösblad, B. (1988). The integration of sensory information in the development of precise manual pointing. *Neuropsychologia, 26*, 805–821.

Wallace, S. A., Newell, K. M., & Wade, M. G. (1978). Decision and response times as a function of movement difficulty in preschool children. *Child Development, 49*, 509–512.

Wickens, C. D. (1974). Temporal limits of human information processing: A developmental study. *Psychological Bulletin, 81*, 739–755.

Wolpert, D. M., Ghahramani, Z., & Flanagan, J. R. (2001). Perspectives and problems in motor learning. *Trends in Cognitive Sciences, 5*, 487–494.

Wolpert, D. M., Miall, R. C., & Kawato, M. (1998). Internal models in the cerebellum. *Trends in Cognitive Sciences, 2*, 338–347.

Zanone, P. G. (1990). Tracking with and without target in 6- to 15-year-old boys. *Journal of Motor Behavior, 22*, 225–249.

Child Development, July/August 2008, Volume 79, Number 4, Pages 866–881

Social and Cognitive Correlates of Children's Lying Behavior

Victoria Talwar
McGill University

Kang Lee
University of Toronto

The relation between children's lie-telling and their social and cognitive development was examined. Children (3–8 years) were told not to peek at a toy. Most children peeked and later lied about peeking. Children's subsequent verbal statements were not always consistent with their initial denial and leaked critical information revealing their deceit. Children's conceptual moral understanding of lies, executive functioning, and theory-of-mind understanding were also assessed. Children's initial false denials were related to their first-order belief understanding and their inhibitory control. Children's ability to maintain their lies was related to their second-order belief understanding. Children's lying was related to their moral evaluations. These findings suggest that social and cognitive factors may play an important role in children's lie-telling abilities.

Lying involves a speaker making a false statement with the intention to deceive the recipient (Bok, 1978; Chrisholm & Freehan, 1977; Coleman & Kay, 1981; Lee, 2000). Lie-telling behavior in children has received increased attention in recent years by developmental psychologists for both its theoretical implications in understanding children's social-cognitive development (e.g., Chandler, Fritz, & Hala, 1989; Leekam, 1993; Peskin, 1992; Polak & Harris, 1999; Sodian, 1991) and its practical applications in legal, clinical, and educational settings (e.g., Chagoya & Schkolne, 1986; Goodman et al., 2006; Lyon, 2000; Stouthamer-Loeber, 1986; Strichartz & Burton, 1990; Talwar, Lee, Bala, & Lindsay, 2002, 2004). The majority of existing research on children and lying has examined children's understanding and moral evaluation of lies (e.g., Bussey, 1992, 1999; Lee, Cameron, Xu, Fu, & Board, 1997; Lee & Ross, 1997; Siegal & Peterson, 1996, 1998; for a review, see Lee, 2000). These studies have shown that children show rudimentary conceptual and moral understanding of lying around 3 years of age but take more than a decade to reach maturity (e.g., being able to consider intention when categorizing a statement as a lie and evaluate its moral values). Only a limited number of studies have investigated children's actual lie-telling behavior, most of which have involved preschool children (e.g., Chandler et al., 1989; Lewis, Stanger, & Sullivan, 1989; Peskin, 1992; Polak & Harris, 1999; Talwar & Lee, 2002a, 2002b; Talwar, Murphy, & Lee, 2007; Talwar et al., 2002, 2004). Overall, these studies have found that lie-telling behavior emerges in the preschool years and that young children are able to deceive others early in life.

Children's Lie-Telling Behavior

The most frequently used method to study lying in children (e.g., Lewis et al., 1989; Polak & Harris, 1999; Talwar & Lee, 2002a; Talwar, Gordon, & Lee, 2007) is the temptation resistance paradigm pioneered by Sears, Rau, and Alpert (1965). In this paradigm, children are typically told explicitly by a researcher not to peek at or play with a toy when left alone. Due to children's curiosity and difficulty to resist temptation (hence, the name of the paradigm), most children tend to disobey the researcher's instruction. Upon returning, the researcher asks children whether they have peeked at or played with the toy. Thus, the temptation resistance paradigm creates a situation where children who have transgressed by disobeying an adults' instruction can make a decision either to lie or to tell the truth about their transgression. The advantage of this paradigm is that it elicits spontaneous lies from children (i.e., children are not instructed to lie) to conceal a transgression. More importantly, it mimics the naturalistic conditions in which children tend to lie (DePaulo & Jordan, 1982; Newton, Reddy, & Bull, 2000; Wilson, Smith, & Ross, 2003).

This research was supported by two separate grants from the Social Sciences and Humanities Research Council of Canada to the authors and National Institute of Child Health and Human Development R01 HD047290 and R01 HD048962 to the second author.

Correspondence concerning this article should be addressed to Victoria Talwar, Department of Educational and Counselling Psychology, McGill University, 3700 McTavish, Montreal, Quebec, Canada H3A 1Y2, or to Kang Lee, Institute of Child Study, University of Toronto, 45 Walmer Road, Toronto, Ontario, Canada M5R 2X2. Electronic mail may be sent to victoria.talwar@mcgill.ca or to kang.lee@utoronto.ca.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7904-0005

Observational studies (e.g., Newton et al., 2000; Wilson et al., 2003) have found that the most common and earliest lies children tell tend to conceal misdeeds where they have done something they were not supposed to do. Interestingly, despite the early and common occurrence of these types of lies, children of all ages and adults view them very negatively. For instance, Bussey (1992, 1999) found that children as young as 4 years of age rated lies about misdeeds as being very bad and that the liar would feel guilty for telling such a lie. Furthermore, they rated this type of lie more negatively than other types of lies and even misdeeds themselves.

In a classic study, Lewis et al. (1989) experimentally investigated 3-year-olds' deception to conceal their transgression with the use of the temptation resistance paradigm. Lewis et al. found that of the 33 children tested, 29 peeked and 38% lied about peeking at the toy. Lewis et al. concluded that children as young as 3 years of age are capable of verbally deceiving others. This result was replicated by Talwar and Lee (2002a) who showed that 36% of 3-year-olds lied about their peeking. They also found that unlike the 3-year-olds, the majority of the children between 4 and 7 years of age lied. Polak and Harris (1999) further modified Lewis et al.'s paradigm. They used a permissive condition where children were allowed to play with a toy and a prohibition condition where children were instructed not to touch the toy. Similar results were obtained with the majority of 3- and 5-year-olds lying in the prohibition condition, whereas all children in the permissive condition admitted to their touching the toy. Thus, Polak and Harris concluded that children's denials about their transgression reflected their deliberate attempt to mislead rather than forgetting. These results, along with the results of related research (Chandler et al., 1989; Lewis et al., 1989; Peskin, 1992; Talwar & Lee, 2002a), suggest that young children are able to engage in intentional verbal deceptive acts when given the opportunity.

Children's Ability to Lie Successfully

While children begin to tell lies from the preschool years, their abilities to lie successfully appear to develop well into middle childhood. To be successful in deceiving an intended dupe, a lie-teller must be able to not only produce a false statement but also ensure consistency between their initial lie with subsequent statements. Any inconsistencies in one's statements may lead to the detection of one's lies. The ability to maintain consistency between statements during deception is referred to as semantic leakage control in the literature (Talwar & Lee, 2002a). Talwar and Lee (2002a) found that when younger children were asked about the identity of the toy, even though children said they had not peeked at the toy, they often failed to feign ignorance and blurted out the identity of the toy. Talwar and Lee (2002a) showed transcripts of the children's exchange with the experimenter to university students who could easily detect children's lies. However, a developmental trend in children's ability at semantic leakage control was found. Whereas the majority of the children between 3 and 5 years blurted out the name of the toy that they denied having peeked at and thus implicated themselves as having transgressed, about half of the 6- and 7-year-olds feigned ignorance of the toy's identity. As a result, those children who feigned ignorance were not distinguishable from children who did not peek, and they were not detected by adults. This finding suggests that with age children become increasingly capable of maintaining consistency in their subsequent statements when lying.

Social and Cognitive Factors

It has been speculated that younger children may lack the cognitive abilities to be convincing lie-tellers (Talwar & Lee, 2002a), which may account for the developmental differences found in children's lie-telling behavior at different ages. However, there has been very little research on the exact social and cognitive factors that may play a role in children's ability to lie successfully. The limited existing research appears to suggest that children's theory-of-mind understanding, executive functioning, and their conceptual and moral understanding of lies may be related to their lie-telling abilities.

Theory-of-mind understanding. There exist two hypotheses regarding the relationship between theory-of-mind understanding and lie-telling. The first hypothesis suggests a relationship between children's lie-telling and their first-order belief understanding (Chandler et al., 1989; Polak & Harris, 1999), herein referred to as ToM_1 Hypothesis. The basis for this hypothesis is that telling a lie successfully requires deliberately creating a false belief in the mind of another. Acts of deception such as lying have been identified as early indicators of a child's understanding of belief and false belief (Chandler et al., 1989; Peskin, 1992; Ruffman, Olson, Ash, & Keenan, 1993; Sodian, 1991). Chandler et al. (1989) found that children as young as 3 years of age would intentionally create a false belief in another by withholding genuine information or planting false information. This conclusion remains controversial (Sodian, 1991;

Sodian, Taylor, Harris, & Perner, 1991). Further evidence of a relation between children's lying and theory-of-mind understanding comes from Polak and Harris (1999) who found that 3- and 5-year-olds' false belief understanding was related to their false denials about having played with a toy. However, their false belief understanding was not related to the lie-tellers' feigning ignorance in follow-up questions (i.e., failing to conceal their transgression and thus revealing the fact that their initial denial was false).

The second hypothesis posits that there is a relationship between children's abilities to maintain their lies and their second-order belief understanding (herein referred to the ToM_2 Hypothesis). Previous research has shown that second-order mental state understanding (e.g., Peter knows that Sally thinks it is raining) begins to emerge only around 6 years of age and undergoes steady development well into adolescence (Sullivan, Zaitchik, & Tager-Flusberg, 1994; Wimmer & Perner, 1983). Further, Banerjee and Yuill (1999) found that children who passed second-order belief tests were more likely to suggest story protagonists make false claims so as to present themselves in a positive light to others. Based on this evidence, Polak and Harris (1999) and Talwar and Lee (2002a) hypothesized, in the context of the aforementioned temptation resistance paradigm, that older children would be more likely to feign ignorance in follow-up questions and their successful deception would be linked to their performance on second-order belief tasks. They suggested that children's false denials of peeking only require the child to represent a belief that is different from the true state of affairs. Therefore, to falsely deny peeking requires only a first-order false belief understanding. However, to sustain the lie that one had not peeked, when asked the identity of the object they were told not to peek at, children had to infer what belief they ought to have, given the initial denial. Thus, Polak and Harris predicted that older children would be more likely to feign ignorance in follow-up questions than younger children and that success in doing so would be linked to performance on second-order false belief tasks. Indirect support for the ToM_2 Hypothesis comes from the findings of Talwar and Lee (2002a) demonstrating that 6- and 7-year-olds (who are known to have better second-order false belief understanding) were better at maintaining their lies than younger children. Recently, Talwar, Gordon, et al. (2007) provided direct support for the ToM_2 Hypothesis. They showed that feigning ignorance after having peeked at a toy was indeed significantly related to second-order belief scores among children between 7 and 11 years of age.

Thus, although initial false denials may be related to children's first-order belief understanding (ToM_1 Hypothesis), maintaining a lie in follow-up questions may be related to a child's ability to represent another's beliefs and what the other will infer from any knowledge revealed by the child (ToM_2 Hypothesis; Polak & Harris, 1999; Talwar, Gordon et al., 2007; Talwar & Lee, 2002a). However, these hypotheses have not been concurrently examined. Only one study has directly examined the ToM_1 Hypothesis (Polak & Harris, 1999) with 3- and 5-year-old children, and only one study has examined the ToM_2 Hypothesis with children 7 years and older (Talwar, Gordon, et al., 2007). The development of children's lie-telling behavior between 3 and 8 years of age, when critical changes in theory-of-mind understanding occurs, has not been examined, nor has the relationship between both children's first- and second-order belief understanding and their actual lie-telling behavior. Thus, one of the objectives of the current study was to examine the relationship between first- and second-order belief understanding and lie-telling abilities in children between 3 and 8 years of age.

Executive functioning. Some evidence exists that children's lie-telling behavior also might be related to executive functioning. Executive functioning has been defined as a set of higher order psychological processes involved in goal-oriented behavior under conscious control (Zelazo & Muller, 2002). Executive functioning encompasses a collection of cognitive skills including self-regulation, inhibitory control, planning, attentional flexibility, and strategy employment (Welsh, Pennington, & Groisser, 1991; Zelazo, Carter, Reznick, & Frye, 1997). Executive functioning skills have been shown to emerge in late infancy and develop during the childhood years (Welsh & Pennington, 1988; Zelazo & Muller, 2002), a time when researchers have noted increases in lie-telling skill (e.g., Polak & Harris, 1999; Talwar & Lee, 2002a). In particular, it has been suggested that inhibitory control and working memory may be directly related to children's deception (Carlson & Moses, 2001; Carlson, Moses, & Hix, 1998). Inhibitory control is the ability to suppress interfering thought processes or actions (Carlson, Moses, & Breton, 2002), and working memory is a system for temporarily holding and processing information in the mind (Baddeley, 1986). When lying, children have to suppress the reporting of the transgression that they wish to conceal and represent and utter the false information that differs from reality (Carlson et al., 1998, 2002). Additionally, to maintain their lies, children must inhibit those thoughts and statements that are contrary to their lie and that would reveal their transgression, while maintaining in their memory the contents of their lie. Thus, to tell lies and to lie successfully, children must be able to hold conflicting alternatives in their mind (i.e., what they really did/thought and what they said they did/thought).

Only one study has examined the relation between children's executive functioning and deceptive behavior (Carlson et al., 1998). Carlson et al. (1998) found that preschool children who experienced difficulty with executive functioning tasks, especially those that require a high level of inhibitory control, demonstrated difficulties with physical deception (i.e., pointing). Although Carlson et al. did not explicitly examine lie-telling behavior, their results seem to suggest that children may also have difficulties with lying if they lack advanced executive functioning skills, particularly in terms of inhibitory control and working memory. However, no study has directly tested this hypothesis. The current study aims to address this gap in the literature by examining the relationship between executive functioning and children's lie-telling abilities.

Children's conceptual and moral understanding of lying and truth-telling. Research has demonstrated that a child's conceptual and moral understanding of lie- and truth-telling emerges early in the preschool years and develops rapidly throughout the school years (Bussey, 1992, 1999; Peterson, Peterson, & Seeto, 1983; Piaget, 1932/1965; Siegal & Peterson, 1998; Talwar et al., 2002; see Lee, 2000, for a review). As young as 3 years of age, children already have a rudimentary concept of lies that are told for antisocial purposes and they evaluate such lies negatively. With increased age, children begin to differentiate antisocial lies from honest mistakes, guesses, exaggerations, and eventually sarcasm and irony. Children also gradually take into consideration the social context in which lies are told and the intention of the lie-teller when evaluating lies. Overall, by early adolescence, children's conceptual and moral understanding of lying and truth-telling becomes comparable to adults.

However, there has been limited research to examine the relationship between children's conceptual and moral understanding and their actual lying behavior. Talwar et al. (2002) found no relationship between children's actual lie- and truth-telling behavior and their conceptual and moral understanding of lies. The majority of children who reported that lying to conceal a transgression was bad, could correctly identify such a lie, and recommended others to tell the truth. Nevertheless, most of them told lies to conceal their own transgressions. Another study by Talwar et al. (2004) found a significant but modest correlation between children's conceptual and moral understanding and their lie-telling behavior to conceal a parent's transgression. It should be noted that in both studies, children's conceptual and moral understanding was assessed via tasks commonly used by legal professionals in the court (Bala, Lee, Lindsay, & Talwar, 2001; Lyon, 2000; Talwar et al., 2004). These tasks are typically very brief and only assessed whether children had a minimal understanding of lying and truth-telling. As a result, there was low variability in children's scores that may have obscured a genuine relationship between children's developing conceptual and moral knowledge and their lie-telling behavior. One of the objectives of the current study was to address this issue by using a more comprehensive measure of children's conceptual and moral understanding of lie- and truth-telling.

In summary, there have been few studies that have examined social and cognitive factors contributing to the development of lying or truth-telling. Furthermore, the few studies that have examined the relationship between social and cognitive factors (e.g., theory of mind, executive functioning, children's conceptual and moral judgment) and lying have only examined one factor and none have examined the differential role of these factors on children's lie-telling behavior. The aims of the present research were thus to address these gaps in the literature and to provide an integrated and comprehensive picture of the relationship between children's lie-telling behavior and various social and cognitive factors.

In the current study, to assess their lying behavior, children between 3 and 8 years of age participated in a temptation resistance paradigm similar to that used in the previous studies (Lewis et al., 1989; Talwar & Lee, 2002a; Talwar et al., 2002). Children were told not to peek at a toy while the researcher left the room. Later, children were asked to promise to tell the truth, and then questioned whether they had peeked. Children's conceptual and moral understanding was assessed using stories where children were asked to identify if the protagonist lied or not and rate their statements as good or bad (Bussey, 1992, 1999; Peterson et al., 1983; Siegal & Peterson, 1998). Children also completed three tests of executive functioning (Gerstadt, Hong, & Diamond, 1994; Hala, Hug, & Henderson, 2003; Kochanska, Murray, Jacques, Koenig, & Vandegeest, 1996) and four belief tasks (Hogrefe, Wimmer, & Perner, 1986; Sullivan et al., 1994; Wimmer & Hartl, 1991; Wimmer & Perner, 1983).

Based on existing evidence, we expected that as age increased, children who peeked at the toy would be more likely not only to deny their transgression (initial lie) but also to conceal their transgression when asked about the identity of the toy (i.e., semantic leakage control). For example, they would feign ignorance about the fact that they knew the identity of the toy. Further, it was expected that children's first-order belief understanding would predict children's lie-telling in the temptation resistance paradigm (ToM_1 Hypothesis) and their second-order belief understanding would predict their semantic leakage control (ToM_2 Hypothesis). It was also expected that better inhibitory control and better working memory would be positively associated with children's semantic leakage control to conceal their transgression in their verbal statements after an initial lie. For example, children who can hold information longer in their working memory and who do not respond impulsively would be more likely to feign ignorance about the identity of the toy. Finally, we hypothesized that children who had high conceptual and moral understanding of lie- and truth-telling (e.g., knowing the difference between truth and lies and its moral negativity) would be less likely to lie in the temptation resistance paradigm.

Method

Participants

Children (N = 150; 80 boys) between 3 and 8 years of age participated in the study (M = 65 months, SD = 12.3, range = 36 to 102 months). There were 77 preschool children (31 girls; M = 49.9, SD = 8.4) and 73 early elementary children (39 girls; M = 85.4, SD = 10.3). Henceforth, the former age group will be referred to as the younger group and the latter the older group. The children were predominately Caucasians and from middle-income families in a medium sized North American city (population: 120,000).

Measures and Procedure

Lie-telling behavior. The present study used a modified version of the temptation resistance paradigm used in previous studies (Lewis et al., 1989; Talwar & Lee, 2002a; Talwar et al., 2002) to examine children's lying behavior. After obtaining parental consent, each child was brought into a room with a researcher. In this room, each child and the researcher played a guessing game. Children were instructed to turn around in their chair so their back was to the researcher while the researcher played a sound from a toy. They were instructed not to break the rule by turning around in their chair to peek at the toy during the game. They could only look when the experimenter said so. Upon hearing the toy sound, the child was asked to guess what the toy was. All the toys were familiar to the children from television programs and stories (e.g., Buzz Light Year, Godzilla, Elmo), and all had familiar accompanying sounds that were clues to their identity. Throughout the game, children were repeatedly told of the no peeking rule. After the child correctly guessed the identity of two toys, the researcher was called out of the room. The researcher told the child she had to leave for a minute and that the third final toy would be left on the table with the sound playing. The third toy was a Barney doll and the sound that accompanied it was unrelated to the doll and came from a greeting card. Thus, the child could not guess correctly the identity of the toy based upon the sound that he/she heard. Prior to the researcher leaving the room, the child was told not to turn around and peek at the toy while the researcher was out of the room. The researcher then left the room for 1 min. A hidden video camera recorded the child's behavior while the researcher was absent. When the researcher, who was unaware of whether the child had peeked, returned to the room they told the child not to turn around and then covered the toy with a sheet. Once the toy was covered, the child was instructed to turn around in their chair. The researcher asked the child to promise to tell the truth. This procedure was adapted from Talwar et al. (2002). Asking children to promise to tell the truth was a major modification of the typical procedure because existing studies showed that without asking children to promise to tell the truth, the majority of children would lie (e.g., Lewis et al., 1989; Polak & Harris, 1999; Talwar & Lee, 2002a; Talwar et al., 2002). This would result in low variability in lying behavior leading to difficulty in revealing any potential relations between lying and social and cognitive factors. However, when children were asked to promise to tell the truth, close to 50% of the children would tell the truth (Talwar et al., 2002), making such a procedure ideal for examining the relations between lying and social and cognitive factors.

The child was then asked the critical question "When I was gone, did you peek at the toy?" and the follow-up question "Who do you think the toy is?" Children who gave a correct answer (Barney) were asked "How did you know who the toy was?"

Social and cognitive factors. Children returned on a second visit to complete three sets of social and cognitive tasks: conceptual and moral judgment tasks, theory-of-mind tasks, and executive functioning tasks. The order of the tasks was counterbalanced and each child was randomly assigned to one of the three different orders. The whole session lasted on average about 30 min with two short breaks given between tasks.

Concept and moral judgment tasks. To test children's conceptual and moral understanding of lie- and truth-telling, nine short stories were used. These stories involved various situations in which a character may have lied to conceal a transgression, told the truth about it, lied for a friend, made an honest mistake, inadvertently passed on a lie, told a white lie, played a trick, told an exaggeration, followed through on a promise, or failed to follow through on a promise. The stories were adapted from Bussey (1992, 1999), Siegal and Peterson (1998), Peterson et al. (1983), and Talwar and Lee (2002a). An example of one of the stories about lying to conceal a transgression is described below:

> Katy's teacher gave her candy for a special treat but told her not to eat it until after lunch. But, when Katy's teacher went away, she ate the candy before lunch. When her teacher came back and asked Katy if she had eaten the candy Katy said no

After each story was complete, the child was asked a concept classification question and an act evaluation question. The questions were modeled based on previous research that has examined children's conceptual and moral evaluations of lying and truth-telling (Bussey, 1992, 1999; Peterson et al., 1983; Siegal & Peterson, 1998; Talwar & Lee, 2002a; Wimmer, Gruber, & Perner, 1984). The concept classification question was, "Is what Katy in the story said a lie, the truth, or something else?" To score this component, each child was given 1 point for each concept question that he or she answered correctly across all stories, yielding a score out of 9. The act evaluation question was, "Is what Katy said, very very bad, very bad, bad, not good/not bad, good, very good, or very very good?" To help answer this question, children were instructed to indicate their answer using a 7-point Likert scale with stars indicating the positive ratings and Xs indicating the negative ratings. For example, if a child pointed to three stars, it indicated she thought the character's behavior was very very good. Two stars indicated very good, and one star indicated just good. This format was used in the same way for the Xs such that one X indicated bad, two Xs very bad, and three Xs very very bad. If a child believed the behavior to be not good and not bad they were instructed to point to the circle. Prior to giving the ratings, each child was trained on how to use the scale. This scale was adapted from previous studies (Bussey, 1992, 1999; Peterson et al., 1983), which found that preschoolers were able to successfully use this scale to evaluate lies. Children's ratings were then converted to scores for each moral question ranging between −3 to +3.

Theory-of-mind understanding. To test children's theory-of-mind understanding, five tasks were used. These five tasks were adapted from Hogrefe et al. (1986), Sullivan et al. (1994), Wimmer and Hartl (1991), and Wimmer and Perner (1983). These tasks included an unexpected contents first-order false belief task, two unexpected location first-order false belief stories, and two unexpected location second-order belief stories. The stories were acted out in puppet plays and shown to children on video. The order of the stories was counterbalanced between participants.

For the unexpected contents false belief task (Wimmer & Hartl, 1991), each child was presented with a box of Band-Aids in which some unexpected contents (i.e., crayons) were placed. The child was shown the Band-Aid box and asked what they thought was in the box. Following the child's answer, the box was opened to reveal a set of crayons. Upon showing the child that the actual contents, the box was closed. Then, the child was asked two questions about the contents of the box. First, the child was asked, "Before you looked inside, what did you think was in the box?" The child was then introduced to a puppet Max and asked, "What does Max think is in the box?" Each child was given 1 point for correctly answering each question. This procedure gave a possible score of 2.

The two unexpected location first-order belief stories were modeled after Wimmer and Perner (1983). One story involved Mark who puts his chocolate in one location before going out to play. His mother moves the chocolate to a second location while he is away. Each child was asked, "Where will Mark look for his chocolate bar?" The second story was the Sally-Ann story where Ann moves Sally's toy car while she is absent. Each child was asked, "Where does Sally think her toy car is?" For each story, the child received a score of 1 point for correctly attributing a false belief to the protagonist, for a total possible score of 2. Along with the contents false belief task, each child had a total "first-order belief score" out of 4.

The two second-order belief tasks were adapted from Hogrefe et al. (1986) and Sullivan et al. (1994). One story involved two children—John and Emma who encounter an ice cream man at the park. Emma goes home to get money for an ice cream. While she is gone, the ice cream man tells John he is going to the school to sell ice cream. On the way to the school, the ice cream man meets Emma and tells her he is going to the school. Throughout the puppet play, the child participant was asked control questions regarding the characters actions (e.g., "What did the ice cream man say to John?"). All children answered the control questions correctly. Each child was asked the following target questions: "Does John know that Emma knows where the ice cream man is now?" and "Where does John think Emma will go to buy ice cream?" Each child received a score of 1 point for correct answers to each question.

The second story involved two children, Mary and Simon, and their Grandpa. In the story, Grandpa gave the children a piece of chocolate to share. Simon wanted to keep the treat for himself so he hid it while Mary played outside. While Simon was busy hiding the chocolate, Mary was watching him through the window. During this puppet play, the child participant was asked several control questions (e.g., "Where has Simon put the chocolate?"). All children answered the control questions correctly. Each child was asked the following target questions: "Does Simon know that Mary knows where chocolate is now?" and "Where does Simon think Mary will look for the chocolate?" Each child received a score of 1 point for correct answers to each question. The total score for the two second-order belief stories was out of 4, which is henceforth referred to as the "second-order belief score."

Executive functioning. To test children's executive functioning, three tasks were used. One task (the whispers task) was a measure of inhibitory control (Kochanska et al., 1996), one (the Stroop task) was a measure of inhibitory control involving inhibition and working memory (Gerstadt et al., 1994; see Carlson & Moses, 2001, for discussion of task) and one (the six-box scramble task) was a measure of working memory (Hala et al., 2003). The order of the tasks was counterbalanced.

In the whispers task, the child was presented with 10 pictures picked out with the aid of their parent. These pictures consisted of story characters that children were familiar with and others that they were not. Child participants were told they would see a series of pictures, and after each picture, they had to tell the researcher who they thought was in the picture. The one rule that children were given for completing this task was that they were to whisper all their answers. To ensure that all the children understood how to whisper, they were asked to whisper their name and their age. It was recorded if the child responded with a whisper, a normal voice, a mixed voice, or a shout. Each child was given 3 points for correctly answering in a whisper, 2 points for answering in a normal voice, 1 point for answering in a mixed voice, and 0 points for answering in a shout. This yielded a possible score of 30.

For the Stroop task, each child was instructed that they would be playing a game of opposites. The child was shown a picture of a moon and asked what it was to ensure the child was familiar with the object. The child was then instructed to say "day or sun" each time the researcher showed them a picture of the moon. The child was then showed a picture of a sun. After ensuring familiarity with the object, the child was instructed to say "night or moon" each time the researcher showed them a picture of the sun. Two practice trials were performed to test the child's understanding of the rules. Following the practice trials, the child was shown 16 pictures, 8 of the sun and 8 of the moon. These pictures were randomly ordered to minimize predictability. Each child was given 1 point for answering correctly and 0 points for answering incorrectly with a total score out of 16.

The final executive functioning task was the six-box scramble task. Six boxes, each a different color, were placed on a plank in front of children. Each child was told that in each box the researcher had placed a sticker and that it was their job to find them. The child was told that he or she was only allowed to select one box at a time. After each selection, the boxes were scrambled and the child was allowed to choose again until he or she found all six stickers. The child was allowed to keep the stickers once he or she found them. To score this task, the number of trials needed for the children to find all stickers were added up and then subtracted from 15 (the maximum number of trials allowed).

Results

To examine children's lie-telling behavior and the relation to different social-cognitive measures, a series of analyses were conducted. First, descriptive statistical results regarding children's peeking and lie-telling behavior were obtained. Second, to examine differences between those children who lied and those who did not in their social and cognitive factor scores, analyses of variance (ANOVAs) were conducted. Finally, to examine what social and cognitive factors predicted children's behavior, hierarchical regression analyses were performed.

Peeking behavior. Overall, 82% of the children (123) peeked at the toy in the experimenter's absence, and on average, children peeked 11 s after the experimenter left the room (SD = 13.2), with half of the children peeking within 6 s. Regression analysis revealed no significant age and sex of child effects for peeking behavior (see below for details).

Initial lie. Of the 123 children who peeked, 79 (64%) children lied about their transgression, whereas 44 children (36%) confessed. Regression analysis revealed no significant age and sex of child effects (see below for details). All 27 children who refrained from peeking stated that they had not peeked.

Lie-teller's responses to the follow-up question (semantic leakage control). Of the 79 children who lied about peeking, 57 (72%) children gave the correct answer to the follow-up question, "What do you think the toy is?" whereas all 44 children who confessed to peeking gave the correct answer, $\chi^2(1, N = 123) = 7.18, p < .01$. Regression analysis revealed no significant age and sex of child effects (see below for details). All 27 children who refrained from peeking gave incorrect answers. Thus, it seems that few lie-tellers concealed their knowledge of the identity of the toy that they had peeked at. Most were unable to maintain their lie in answer to the follow-up question.

Children's answers to "How did you know what the toy was?" when they gave the correct answer were analyzed to examine further children's ability to maintain their initial lie. Children's answers were divided into two categories: plausible explanations that attempted to conceal their peeking (e.g., "I heard the music before and knew it was Barney," "My sister has the video and I hear that music on it," "I knew the song was Barney's") and explanations that revealed their peeking behavior or did not explain the correct answer (e.g., "I saw purple," "It looked like Barney," "I don't know") with 97% intercoder reliability. Of the 57 children who said the toy was Barney, 28 (49%) gave plausible answers, whereas 29 either said "I don't know" or implicated themselves (e.g., "I saw purple"). Regression analysis revealed a significant effect of age but no sex of child effect (see below for details). With increased age, children became better at maintaining their lies and giving plausible explanations that concealed their transgression. To illustrate this age effect, we dichotomized children's ages into a younger group (77 preschoolers: $M = 49.9$, $SD = 8.4$; 31 girls) and an older group (73 early elementary school children: $M = 85.4$, $SD = 10.3$; 39 girls): Seventy-seven percent of the older children gave plausible explanations, whereas only 29% of the younger children gave plausible explanations.

Univariate Analyses: Social and Cognitive Measures Related to Children's Lie-Telling Behavior

To further examine the differences between lie-tellers, confessors, and children who did not peek (nonpeekers) in terms of each social and cognitive factors (i.e., theory-of-mind understanding, executive function, and conceptual moral understanding), we conducted a 3 (type of child: liar, confessor, nonpeekers) × 2 (gender) × age (continuous variable) ANOVA on each social and cognitive factor. Age was treated as a continuous variable in these analyses to help increase the power of the analysis because only 1 *df* was used. The reason that individual ANOVAs were conducted on each social or cognitive measure was to ascertain whether our initial a priori hypothesis regarding each measure was correct.

Whispers task. There was a significant effect for age, $F(1, 149) = 16.36, p < .001, \eta^2 = .13$. To illustrate the significant age effect for this task as well as the other social-cognitive tasks (see below), we again dichotomized children's ages into two age groups: the younger children (the preschoolers) and the older children (the early elementary school children). As shown in Table 1, the older children's scores on the whispers task were significantly greater than the younger children's scores. No other factors were significant. There were no differences between lie-tellers, confessors, and nonpeekers.

Stroop task. There was a significant effect for age, $F(1, 149) = 41.28, p < .001, \eta^2 = .22$. Children's scores on the Stroop task increased with age (Table 1). There was also a significant effect for type of child, $F(2, 149) = 5.32, p < .01, \eta^2 = .07$. Bonferroni post hoc analyses revealed that confessors had significantly lower scores ($M = 10.68$, $SD = 4.47$) than lie-tellers ($M = 13.04$, $SD = 3.80$, $p < .01$) and nonpeekers ($M = 13.48$, $SD = 2.87$, $p < .05$). There were no significant differences between the two latter groups.

Six-box scramble task. There was a significant effect for age, $F(1, 149) = 7.27, p < .01, \eta^2 = .05$. Children's scores increased with age (Table 1). There was also

Table 1
Mean (SDs) Scores of Executive Functioning, Theory of Mind, and Conceptual Moral Knowledge by Age Group

	Younger children	Older children
Executive function		
Whispers	25.36 (6.14)	28.63 (2.27)
Stroop	11.09 (4.54)	13.83 (2.71)
Six-box scramble	7.58 (2.23)	8.45 (1.38)
Theory of mind		
First-order belief	1.88 (1.43)	3.40 (1.09)
Second-order belief	2.08 (0.99)	2.99 (0.54)
Conceptual moral judgments		
Conceptual knowledge	7.23 (2.03)	9.60 (2.50)
Moral evaluations		
Factuality	0.24 (1.15)	−0.25 (0.74)
Motivation	−0.15 (1.04)	0.16 (0.94)
Promise	0.11 (0.94)	−0.11 (1.05)

a significant effect for sex of child, $F(2, 149) = 7.27, p < .01, \eta^2 = .05$. Boys had lower scores on the six-box scramble ($M = 7.58, SD = 2.28$) than girls ($M = 8.51, SD = 1.22$). There were no differences between lie-tellers, confessors, and nonpeekers.

First-order belief. There was a significant effect for age, $F(1, 149) = 71.67, p < .001, \eta^2 = .38$. Children's first-order belief scores increased with age (Table 1). There was also a significant effect for type of child, $F(2, 149) = 4.79, p = .01, \eta^2 = .07$. Bonferroni post hoc analyses with lie-tellers as the reference group revealed that lie-tellers had significantly higher scores ($M = 3.04, SD = 1.28$) than confessors ($M = 1.93, SD = 1.50, p < .05$) or nonpeekers ($M = 2.52, SD = 1.60, p < .05$). Thus, it appears that ToM_1 Hypothesis was confirmed with children who initially lied having higher scores than those who confessed or did not peek.

Second-order belief. There was a significant effect for age, $F(1, 149) = 47.56, p < .001, \eta^2 = .25$. Children's second-order belief scores increased with age (Table 1). There were no other significant effects.

Children's conceptual knowledge of truthful and untruthful statements. There was a significant effect for age, $F(1, 149) = 91.89, p < .001, \eta^2 = .2$. Children's ability to correctly classify lies increased with age (Table 1). There were no other significant effects.

Children's moral judgments of truthful and untruthful statements. A factor analysis using the principal component method of extraction with varimax rotation was performed on children's ratings of the different stories (i.e., their answers to the act evaluation questions). This analysis yielded three factors that accounted for 54% of the variance (see Table 2). The first factor, labeled Factuality, contained evaluations of three stories (factor loadings above 0.4) including a trickery, white lie, and lying to conceal a transgression (eigenvalue = 2.29, 26% of the variance accounted for). The second factor, labeled Motivation, contained evaluations of four stories including lies for a friend, exaggerations, inadvertent lies, and mistakes (eigenvalue = 1.44, 16% of the variance accounted for). The third factor, labeled Promise, contained evaluations of two stories regarding promising and failure to keep a promise (eigenvalue = 1.15, 13% of the variance accounted for). For subsequent analyses, factor scores representing the three factors were derived.

Table 2
Factor Loading of Moral Judgments of Verbal Statements

	Factor		
Moral evaluation	Factuality	Motivation	Promise
Fail to follow through on a promise	.09	.15	−.87
Lie on behalf of a friend	.36	.46	−.03
White lie situation	.70	−.01	−.03
Lie about a transgression	.73	.27	−.02
Follow through on a promise	.15	.23	.78
Exaggeration	.30	.57	−.02
Trickery	.70	.12	.13
Inadvertently pass on a lie	.03	.63	.27
Mistake	−.10	.74	−.07

Three separate ANOVAs were performed for each of the three factor scores: Factuality, Motivation, and Promise with age as a continuous variable and type of child as a categorical variable. For the Factuality factor, there was a significant effect for age, $F(1, 149) = 4.92, p < .05, \eta^2 = .04$. Children's Factuality factor scores decreased with age (Table 1). There was also a significant effect for type of child, $F(2, 149) = 3.10, p < .05, \eta^2 = .04$. Bonferroni post hoc analyses revealed that confessors had higher Factuality factor scores ($M = 0.32, SD = 0.83$) than lie-tellers ($M = -0.17, SD = 1.05, p < .05$) or nonpeekers ($M = -0.02, SD = 0.99, p < .05$). For the Motivation factor, there was a significant effect for age, $F(1, 149) = 9.46, p < .01, \eta^2 = .06$. Children's Motivation factor scores increased with age (Table 1). There were no other significant effects. For the Promise factor, there was a significant effect for age, $F(1, 149) = 4.65, p < .05, \eta^2 = .03$. Children's Promise factor scores decreased with age (Table 1). There was also a significant effect for type of child, $F(2, 149) = 3.48, p < .05, \eta^2 = .05$. Bonferroni post hoc analyses revealed that nonpeekers had higher Promise factor scores ($M = .54, SD = 1.09$) than confessors ($M = -0.08, SD = .81, p < .05$) or lie-tellers ($M = -0.1, SD = 1.03, p < .05$).

Multivariate Analyses: Social and Cognitive Measures Relationship to Children's Behavior

Given the presence of several significant correlations among the independent variables shown in Table 3, a series of hierarchical regressions were conducted on children's behavior in the temptation resistance paradigm to examine whether social and cognitive factors were together or individually predictive of children's peeking (i.e., whether they peeked or did not peek), tendency to lie (i.e., whether they lied or told the truth), and ability to maintain their lies (i.e., whether they were able to feign ignorance after having told a lie). Based on previous research, it was expected that children's lie-telling behavior would be predicted by their theory of mind and executive function abilities.

Peeking behavior. A hierarchical logistic regression was conducted with children's peeking behavior

Table 3
Correlation Matrix Among Individual Social and Cognitive Variables

	1	2	3	4	5	6	7
1. Concept	—	—	—	—	—	—	—
2. Inhibitory control	.08	—	—		—	—	—
3. Stroop	.18*	.28**	—		—	—	—
4. Working memory	.21*	.26**	.07		—	—	—
5. False belief	.33**	.36**	.32**	.17*		—	—
6. Factor 1 (Factuality)	.16*	−.17*	−.14	−.04	.28**		—
7. Factor 2 (Motivation)	−.40**	−.11	−.13	−.13	−.25**	.00	
8. Factor 3 (Promise)	.14	.08	.09	.06	.19*	.00	.00

*$p \leq .05$. **$p \leq .01$.

(peeking or not peeking) as the predicted variable and child age (continuous variable) and sex entered in the first step and whispers task score, Stroop task score, six-box scramble task score, first-order and second-order belief score, concept classification score, and three factor scores were entered on a second step. For this and subsequent logistic regression analyses, the independent variables, because they were chosen for theoretical reasons (see Menard, 2002), were first entered as predictors. Additional predictors (i.e., interactions) were added individually to determine whether they would contribute significantly to the model. Significance was assessed by a Block chi-square test (also known as the chi-square difference test). In this test, the retention of each predictor in a model must increase the variability substantially in order to justify using a more complex model. The overall model was not significant, $\chi^2(11, N = 150) = 16.74$, *ns*.

Peek latency. In a hierarchical linear regression analysis with peekers' peek latency scores as the predicted variable and the same predictors described above, the overall model was not significant, $F(11, 111) = 1.07$, *ns*.

Initial lie (tendency to lie). A hierarchical logistic regression was conducted with children's initial tendency to lie or tell the truth as the predicted variable and child age and sex were entered in the first step and whispers task score, Stroop task score, six-box scramble task score, first-order and second-order belief score, concept classification score, and three factor scores were entered on a second step. The dependent variable was children's answers to the question "Did you peek?" (liars vs. confessors). The first model was not significant, $\chi^2(2, N = 123) = 1.51$, *ns*. The second model was significant, $\chi^2(11, N = 123) = 28.02$, $p < .01$. Children's Stroop scores were a significant predictor of lying ($\beta = 0.2$, Wald $= 8.5$, $p < .01$). Children who lied had higher Stroop task scores ($M = 13.04$, $SD = 3.8$) than children who confessed ($M = 10.68$, $SD = 4.47$). Children's first-order belief scores was a significant predictor of lying ($\beta = 0.88$, Wald $= 5.76$, $p < .05$). Children who lied had higher first-order belief scores ($M = 3.04$, $SD = 1.29$) than children who confessed ($M = 1.93$, $SD = 1.50$).

Lie-teller's responses to the follow-up question (semantic leakage control). A logistic regression analysis was conducted with children's responses to the follow-up question "Who do you think it is" (correct vs. incorrect) as the predicted variable. Child age and sex were entered in the first step, whispers task score, Stroop task score, six-box scramble task score, first-order and second-order belief score, concept classification score, and three moral judgment factor scores were entered on a second step (confessors were excluded from the analysis). The first model was not significant, $\chi^2(2, N = 79) = 1.63$, *ns*. The second the model was significant, $\chi^2(11, N = 79) = 23.80$, $p < .05$. Children's stroop scores were a significant predictor of lying ($\beta = -0.41$, Wald $= 3.9$, $p < .05$). Children who gave incorrect answers had higher Stroop task scores ($M = 14.86$, $SD = 1.13$) than children who gave the correct answer ($M = 12.33$, $SD = 4.23$).

A logistic regression was conducted on lie-tellers' explanations for giving a correct answer. Child age and sex were entered in the first step, and whispers task score, Stroop task score, six-box scramble task score, first-order and second-order belief score, concept classification score, and three moral judgment factor scores were entered on a second step (confessors were excluded from the analysis). The overall regression model was significant, $\chi^2(11, N = 57) = 23.42$, $p < .05$. In the first step, children's age significantly predicted children's explanations ($\beta = 0.79$, Wald $= 11.61$, $p < .001$). In the second step after age was partialed out, children's second-order belief scores were a significant predictor of children's explanations ($\beta = 1.47$, Wald $= 6.41$, $p < .05$). Children who gave plausible explanations that concealed their peeking had higher second-order belief scores ($M = 2.9$, $SD = 0.69$) than children who gave the correct answer ($M = 2.1$, $SD = 0.86$). Thus, present results supported the ToM$_2$ Hypothesis with children's second-order belief understanding predicting their ability to maintain their lies and provide plausible explanations that did not reveal their transgression.

Discussion

The current study investigated that children's lie-telling behavior and its relation to their theory of mind, executive functioning, and conceptual and moral knowledge of lie- and truth-telling was examined. Several major findings were obtained.

First, not only did most children peek at the forbidden toy, but also, when asked by an experimenter if they had peeked, 64% denied their transgression and thus lied. These findings are consistent with past research, which has found a strong tendency to lie in children under 7 years of age if they have transgressed (Lewis et al., 1989; Polak & Harris, 1999; Talwar & Lee, 2002a). However, unlike previous studies that did not ask children to promise to tell the truth and found that 3-year-olds were less likely to lie compared to older children, no such difference was found in the present study. Our findings are similar to Talwar et al. (2002) who found no age difference when children were asked to promise to tell the truth. Also like Talwar et al. (2002), we found a similar rate of lying and confessing with more children confessing after promising to tell the truth than that found in other studies when children were not asked to promise to tell the truth. Taking together the present and previous findings, it appears that by 3 years of age and after, the tendency to lie about one's own transgression remains strong throughout preschool and elementary school years. However, asking young children to promise to tell the truth can reduce their tendency to lie.

Second, it was hypothesized that older children would not only deny their transgression but also successfully control semantic leakage by concealing their lie in the follow-up questioning. Overall, lie-tellers did not feign ignorance in their response to the first follow-up question. They correctly named the identity of the toy. However, half of the children appeared to make attempts to conceal their transgression by providing seemingly plausible explanations about why they knew the identity of the toy and the tendency to give plausible explanations increased with age. Younger lie-tellers were more likely to correctly name the identity of the toy after claiming to not having peeked at the toy and then failed to give a plausible explanation for knowing the identity. Thus, it seems that as age increases, older children are more skilled at maintaining their lies in subsequent verbal statements after they have told an initial lie.

Third, it was hypothesized that children's lie-telling behavior would be related to their first- and second-order belief understanding. Specifically, it was expected that children who performed better on first-order belief tasks would be more likely to lie (ToM$_1$ Hypothesis) and children who performed better on second-order belief tasks would be better at maintaining semantic leakage control (ToM$_2$ Hypothesis). The present results support both hypotheses. Consistent with Polak and Harris (1999), children's performance on first-order belief tasks predicted the false denials of their peeking behavior. Thus, children's initial lies may reflect their ability to represent a false belief that is different from their belief about the true state of affairs.

Furthermore, those children who had higher second-order belief scores were also better at concealing their transgression in follow-up questions, thus supporting the ToM$_2$ Hypothesis (Polak & Harris, 1999; Talwar & Lee, 2002a). Our results are consistent with the findings of Talwar et al. (2007) that the second-order belief understanding of elementary school-aged children (7 years and older) is significantly related to their ability to maintain semantic leakage control. However, there exists a noteworthy difference between the present study and that of Talwar et al. (2007). In Talwar et al. (2007), children's second-order belief understanding was significantly related to their ability to feign ignorance. In contrast, in the present study, children's second-order belief understanding was not significantly related to their ability to feign ignorance but rather to their ability to provide plausible explanations to their correct naming of the forbidden toy. This difference, however, does not suggest inconsistencies between the two studies because the oldest children (7- and 8-year-olds) in the present study, like the youngest children (7- and 8-year-olds) in the Talwar et al. study, tended to blurt out the correct name of the forbidden toy. Thus, seeming inconsistencies in correlational results may reflect a developmental progression in children's ability at semantic leakage control, from being unable to control the leakage at all during the preschool years, to merely explaining away their leakage with plausible answers in the early elementary school years, to exercise greater leakage control by feigning ignorance altogether. It should be noted that future studies are needed to examine these differences and the relationship between children's theory-of-mind understanding and other types of lies (e.g., high-stakes lies where there are negative consequences to others, prosocial lies told to benefit others). It may be that when children must consider the consequences to others when lying, sharper developmental differences will be found in relation to their theory-of-mind understanding.

Nevertheless, these findings along with previous research begin to provide a developmental picture of children's lying to conceal their own transgression. Children's lying appears to progress through three levels (see also Polak & Harris, 1999; Talwar et al., 2007). First, children's "primary lies" begin around 2–3 years of age when children are first able to deliberately make factually untrue statements. Although it is still unclear whether such statements are a form of word play, wish fulfillment, or genuine deception (i.e., statements made with an intent to instill false belief into the mind of the recipient), children's first falsehoods are often linked to situations of rule violations and children's attempt to avoid incrimination, protect self-interests, or present themselves in a more positive light (Newton et al., 2000; Wilson et al., 2003). Given the fact that genuine lies told by children in later childhood tend to serve similar functions, such early falsehood may be a rudimentary form of intentional verbal deception. However, at this age, children's lies are still infrequent (Newton et al., 2000; Wilson et al., 2003) with approximately half of the 3-year-olds lying about their transgressions, whereas the remainder tend to be honest and confess their transgression when asked by adults (e.g., Lewis et al., 1989; Polak & Harris, 1999; Talwar & Lee, 2002a).

The second level, "secondary lies," reflects a significant shift that takes place between 3 and 4 years of age (Chandler et al., 1989; Peskin, 1992; Polak & Harris, 1999). At and after 4 years of age, the majority of children will readily tell a lie to conceal their own transgression. The results of the present study and previous studies show that children's acquisition of first-order belief understanding may play an important role in children's progression from the first to the second level. Perhaps also related to the development of first-order belief understanding, children are able to regulate successfully their nonverbal behaviors to appear honest (Talwar & Lee, 2002a). However, many children at the secondary level appear to have difficulty at semantic leakage control. Their subsequent statements following an initial false statement tend to be inconsistent with the initial lie and thus make their deception readily detected by naïve adults (Talwar & Lee, 2002a). The third level, "tertiary lies," emerges around 7–8 years of age. At this level, children become gradually more and more sophisticated at semantic leakage control. Children will tell a deliberate lie while ensuring that their subsequent statements do not contradict the initial lie and thus make their statements difficult to distinguish from statements made by a nonliar. As found in the present study, children's second-order belief understanding may play an important role in the transition from the secondary to the tertiary level. This is perhaps because such understanding "allows intentional social coordination to occur" (Perner, 1988, p. 272) such that children can reason about complex interactions between mental states involved in sustaining a lie and act appropriately.

Fourth, the present study found a significant relation between children's executive functioning and their lying behavior. Children with higher Stroop task scores were more likely to lie. When children were asked if they peeked, children had to suppress the reporting of the transgression that they wished to conceal and represent and utter the false information that differs from reality. The inhibitory control that is needed to tell such a lie may be the same executive functioning skills that are involved in performing the Stroop task. Our findings are generally in line with Carlson et al. (1998) who found that children who experienced difficulty with executive functioning tasks, especially those that require a high level of inhibitory control, demonstrated difficulties with deception tasks. The Stroop task also involves working memory (see Carlson & Moses, 2001). It is thus possible that working memory may also play a role in children's decision to lie. Telling a lie may require the dual ability to remember the rule being violated and inhibit reporting of the transgression that they wish to conceal. However, it should be noted that two other executive functioning tasks (one of inhibitory control and one of working memory) were not found to be related to lying. Thus, further investigation of the relations between lying and both inhibitory control and working memory is required to examine the differential impact of these abilities by using tasks where inhibitory control and working memory can be readily isolated. Nevertheless, the present findings suggest that the development of children's lie-telling abilities is not only related to children's theory-of-mind abilities but also to their executive functioning skills. This finding may not be surprising because research has consistently shown a significant relationship between children's theory-of-mind understanding and their executive functioning (e.g., Carlson et al. 2002; Hughes, 1998; Perner, Lang, & Kloo, 2002). Interestingly, dual executive demand tasks like the

Stroop task used in this study have been found to be more strongly predictive of theory-of-mind capabilities than working memory or inhibitory tasks alone (e.g., Carlson & Moses, 2001; Hala et al., 2003). Given this finding, it may be that the combination of inhibitory control and working memory may be crucial not only for ToM reasoning but also for lying, which requires children to put their theory-of-mind knowledge into action.

The fifth major finding of the present study is the lack of a significant relation between children conceptual understanding of lie- and truth-telling and children's actual behavior. Consistent with previous studies, in the current study, children's concept classification scores increased with age (Bussey, 1992, 1999; Lee, 2000; Peterson et al., 1983; Siegal & Peterson, 1998; Talwar et al., 2002). Although a much more comprehensive measure of children's conceptual knowledge was used in the present study, the findings were similar to the two other studies that have examined the relationship between children's conceptual understanding and their actual lying behavior (Talwar et al., 2002, 2004). Perhaps this is not surprising given that adults know what a lie is but still lie on a day-to-day basis (DePaulo & Kashy, 1998). Taken together, these findings suggest that children's ability to classify truth and lies does not prevent them from telling lies to conceal their transgressions. Future studies should examine if this is also true for children's lie-telling behavior in other situations such as lies told for prosocial reasons (e.g., white lies).

Similar to previous studies (e.g., Bussey, 1992, 1999; Piaget, 1932/1965; Siegal & Peterson, 1998), children's age was significantly related to their evaluations of the moral stories. When making moral judgments, younger children were more likely to attend to factors of factuality and promising than older children. However, they were less likely to take motivation into consideration when compared to older children. Thus, younger children paid more attention to the factuality of a statement and the adherence or violation of rules (e.g., promise keeping or breaking) to make their moral evaluations, whereas older children considered the character's intention to deceive itself to make their evaluations. These findings are in keeping with previous studies that have found that although children as young as 4 years of age can make basic distinctions between lies and truth, their moral understanding of lies develops over time with younger children being more influenced by factuality of statements and external factors, whereas older children are more influenced by intentions and internal factors (e.g., Bussey, 1992, 1999; Peterson et al., 1983; Piaget, 1932/1965). Therefore, to fully understand children's emerging understanding of lies, it is important to examine not only their classification abilities but also their evaluations of truthful and untruthful statements.

More importantly, the sixth major finding of the present study showed children's lying behavior to be related to their moral evaluations. These findings are different from previous studies (Talwar et al., 2002, 2004) that found no or limited relation between moral understanding and behavior, perhaps due to the fact that the current study used a more comprehensive measure of children's moral understanding. More specifically, the confessors had higher Factuality scores than lie-tellers. In other words, children who admitted their transgression were more likely to value truthfulness and give it higher ratings regardless of the situation. In contrast, children who chose to lie tended not to have stringent views about the need to be truthful. These results suggest that children who hold more relativist views about the moral implication of lying might be more inclined to tell lies, whereas those who held more stringent moral views about lying are more likely confess. Another noteworthy finding is that in comparison with the liars and confessors, nonpeekers gave the most positive ratings for stories where the protagonist kept a promise and the most negative ratings when the story protagonist failed to keep a promise. One possible explanation is that nonpeekers were the most concerned about rules and adherence to them. They might have taken the experimenter's instruction about not peeking at the forbidden toy more seriously than the lie-tellers and confessors, and this concern was strong enough to motivate them to resist the high-level temptation to peek in the current procedure. Thus, although children's ability to classify whether a statement was a lie was not related to their behavior, their perceptions of the acceptability of such statements were significantly related to their behavior.

A few caveats to our findings are warranted. First, the present study involved children' disobeying an adult's instruction and lying to them about their transgression. Although previous research suggests that this kind of lie is common among children told to parents and teachers and they also considered it to be a serious moral violation (Bussey, 1992, 1999; DePaulo & Jordan, 1982; Newton et al., 2000; Stouthamer-Loeber, 1986; Wilson et al., 2003), the violation itself is not severe and hence may be of low stake in nature. When the stakes are high or consequences of transgressions are more serious, children's behavior as well as its relation to moral judgment may change. Given the obvious ethical challenges of creating such situations, this type of lie has not yet been systematically examined. Nevertheless, research with adults suggests that in high-stakes situations, we are less likely to conform to our moral knowledge and standards and more likely to act out of self-oriented motivation (Batson & Thompson, 2001). Second, the present study was the first study to find a relation between children's moral understanding and their lie-telling behavior with the use of more extensive measures of children's moral understanding than those used in previous studies that failed to find a significant relation. Despite this methodological improvement, it is possible that a more refined assessment may yield an even stronger relationship. Furthermore, because we only used one behavioral measure, it is unclear whether children's moral understanding is also related to other types of lying such as white-lie-telling. Future research with multiple behavioral and moral understanding measures will provide necessary evidence to elucidate how children's behaviors are influenced by their moral knowledge about lying and honesty and vice versa.

In summary, current findings demonstrate that the majority of children between 3 and 8 years of age will lie to conceal their transgressions and their ability to maintain these lies increases with age. Children's ability at semantic leakage control increased not only with age but also with increased cognitive sophistication. Lie-telling behavior was related to both children's theory of mind and executive functioning abilities. Although children's deception is often considered to be problematic, the current study's results suggest that lying is associated positively with children's cognitive development in terms of their understanding of others' minds and executive functioning. The present study serves only as the first step in understanding possible contributing factors to the development of children's deception, focusing only on the contribution of three major social and cognitive factors to children's lie-telling behavior. Other factors that may influence children's deceptive abilities need to be examined in future studies. For instance, children's general intellectual ability, parenting styles, disciplinary styles, and cultural contexts may be also related to the development of lie-telling behavior (Achenbach & Edelbrock, 1981; Cole & Mitchell, 1998; Crossman & Lewis, 2006; Lee et al., 1997; Lewis, 1993; Stouthamer-Loeber & Loeber, 1986).

References

Achenbach, T. M., & Edelbrock, C. S. (1981). Behavioral problems and competencies reported by parents of normal and disturbed children aged four through sixteen. *Monographs for the Society for Research in Child Development, 46* (Serial No. 188).

Baddeley, A. D. (1986). *Working memory.* Oxford, UK: Clarendon Press.

Bala, N., Lee, K., Lindsay, R. C. L., & Talwar, V. (2001). A legal and psychological critique of the present approach to the assessment of the competence of child witnesses. *Osgoode Hall Law Journal, 38,* 409–451.

Banerjee, R., & Yuill, N. (1999). Children's understanding of self-presentational display rules: Associations with mental-state understanding. *British Journal of Developmental Psychology, 17,* 111–124.

Batson, C. D., & Thompson, E. R. (2001). Why don't moral people act morally? Motivational considerations. *Current Directions in Psychological Science, 10,* 54–57.

Bok, S. (1978). *Lying: Moral choice in public and private life.* New York: Pantheon Books.

Bussey, K. (1992). Lying and truthfulness: Children's definitions, standards, and evaluative reactions. *Child Development, 63,* 129–137.

Bussey, K. (1999). Children's categorization and evaluation of different types of lies and truths. *Child Development, 70,* 1338–1347.

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development, 72,* 1032–1053.

Carlson, S. M., Moses, L. J., & Breton, C. (2002). How specific is the relation between executive functioning and theory of mind? Contribution of inhibitory control and working memory. *Infant and Child Development, 11,* 73–92.

Carlson, S. M., Moses, L. J., & Hix, H. R. (1998). The role of inhibitory processes in young children's difficulties with deception and false belief. *Child Development, 69,* 672–691.

Chagoya, L., & Schkolne, T. (1986). Children who lie: A review of literature. *Canadian Journal of Psychiatry, 31,* 665–669.

Chandler, M. J., Fritz, A. S., & Hala, S. M. (1989). Small-scale deceit: Deception as a marker of two-, three-, and four-year-olds' early theories of mind. *Child Development, 60,* 1263–1277.

Chrisholm, R. M., & Freehan, T. D. (1977). The intent to deceive. *Journal of Philosophy, 75,* 143–159.

Cole, K., & Mitchell, P. (1998). Family background in relation to deceptive ability and understanding of the mind. *Social Development, 7,* 181–197.

Coleman, L., & Kay, P. (1981). Prototype semantics: The English word lie. *Language, 1,* 26–44.

Crossman, A. M., & Lewis, M. (2006). Adult's ability to detect children's lying. *Behavioral Sciences & Law, 24,* 703–715.

DePaulo, B. M., & Jordan, A. (1982). Age changes in deceiving and detecting deceit. In R. S. Feldman (Ed.), *Development of non-verbal behavior in children* (pp. 151–179). New York: Springer-Verlag.

DePaulo, B. M., & Kashy, D. A. (1998). Everyday lies in close and casual relationships. *Journal of Personality and Social Psychology, 74,* 63–79.

Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: Performance of children 3–7 years old on a Stroop-like day/night test. *Cognition, 53,* 129–153.

Goodman, G. S., Myers, J. E. B., Qin, J., Quas, J. A., Castelli, P., Redlich, A. D., et al. (2006). Hearsay versus children's testimony: Effects of truthful and deceptive statements on juror's decisions. *Law and Human Behavior, 30,* 363–401.

Hala, S., Hug, S., & Henderson, A. (2003). Executive functioning and theory of mind in preschool children: Two tasks are harder than one. *Journal of Cognition and Development, 4,* 275–298.

Hogrefe, G. J., Wimmer, H., & Perner, J. (1986). Ignorance versus false belief. A developmental lag in attribution of epistemic states. *Child Development, 53,* 567–582.

Hughes, C. (1998). Finding your marbles: Does preschoolers strategic behaviour predict later understanding of mind? *Developmental Psychology, 34,* 1326–1339.

Kochanska, G., Murray, K., Jacques, T. Y., Koenig, A. L., & Vandegeest, K. A. (1996). Inhibitory control in young children and its role in emerging internalization. *Child Development, 67,* 490–507.

Lee, K. (2000). The development of lying: How children do deceptive things with words. In J. W. Astington (Ed.), *Minds in the making* (pp. 177–196). Oxford, England: Blackwell.

Lee, K., Cameron, C. A., Xu, F., Fu, G., & Board, J. (1997). Chinese and Canadian children's evaluations of lying and truth-telling. *Child Development, 64,* 924–934.

Lee, K., & Ross, H. (1997). The concept of lying in adolescents and young adults: Testing Sweester's folkloristic model. *Merrill-Palmer Quarterly, 43,* 255–270.

Leekam, S. (1993). Children's understanding of mind. In M. Bennett (Ed.), *The development of social cognition: The child as psychologist* (pp. 26–61). New York: Guilford Press.

Lewis, M. (1993). The development of deception. In M. Lewis & C. Saarni (Eds.), *Lying and deception in everyday life* (pp. 90–105). New York: Guilford Press.

Lewis, M., Stanger, C., & Sullivan, M. (1989). Deception in 3-year-olds. *Developmental Psychology, 25,* 439–443.

Lyon, T. D. (2000). Child witnesses and the oath: Empirical evidence. *Southern California Law Review, 73,* 1017–1074.

Menard, S. (2002). *Quantitative applications in the social sciences: No. 106. Applied logistic regression analysis* (2nd ed.). Thousand Oaks, CA: Sage.

Newton, P., Reddy, V., & Bull, R. (2000). Children's everyday deception and performance on false-belief tasks. *British Journal of Developmental Psychology, 18,* 297–317.

Perner, J. (1988). Higher-order beliefs and intentions in children's understanding of social interaction. In J. W. Astington, P. L. Harris, & D. R. Olson (Eds.), *Developing theories of mind* (pp. 271–294). New York: Cambridge University Press.

Perner, J., Lang, B., & Kloo, D. (2002). Theory of mind and self-control: More than a common problem of inhibition. *Child Development, 73,* 752–767.

Peskin, J. (1992). Ruse and representations: On children's ability to conceal information. *Developmental Psychology, 28,* 84–89.

Peterson, C. C., Peterson, J. L., & Seeto, D. (1983). Developmental changes in ideas about lying. *Child Development, 54,* 1529–1535.

Piaget, J. (1965). *The moral judgment of the child* (M. Gabain, Trans.). New York: Free Press. (Original work published 1932)

Polak, A., & Harris, P. L. (1999). Deception by young children following noncompliance. *Developmental Psychology, 35,* 561–568.

Ruffman, T., Olson, D. R., Ash, T., & Keenan, T. (1993). The ABC's of deception: Do young children understand deception in the same way as adults? *Developmental Psychology, 29,* 74–87.

Sears, R., Rau, L., & Alpert, R. (1965). *Identification and child rearing.* New York: Wiley.

Siegal, M., & Peterson, C. C. (1996). Breaking the mold: A fresh look at children's understanding of questions about lies and mistakes. *Developmental Psychology, 32,* 322–334.

Siegal, M., & Peterson, C. C. (1998). Preschoolers' understanding of lies and innocent and negligent mistakes. *Developmental Psychology, 34,* 332–341.

Sodian, B. (1991). The development of deception in young children. *British Journal of Developmental Psychology, 9,* 173–188.

Sodian, B., Taylor, C., Harris, P., & Perner, J. (1991). Early deception and the child's theory of mind: False trails and genuine markers. *Child Development, 62,* 468–483.

Stouthamer-Loeber, M. (1986). Lying as a problem behavior in children: A review. *Clinical Psychology Review, 6,* 267–289.

Stouthamer-Loeber, M., & Loeber, R. (1986). Boys who lie. *Journal of Abnormal Child Psychology, 14,* 551–564.

Strichartz, A. F., & Burton, R. V. (1990). Lies and truth: A study of the development of the concept. *Child Development, 61,* 211–220.

Sullivan, K., Zaitchik, D., & Tager-Flusberg, H. (1994). Preschoolers can attribute second-order beliefs. *Developmental Psychology, 30,* 395–402.

Talwar, V., Gordon, H., & Lee, K. (2007). Lying in the elementary school: Verbal deception and its relation to second-order belief understanding. *Developmental Psychology, 43,* 804–810.

Talwar, V., & Lee, K. (2002a). Development of lying to conceal a transgression: Children's control of expressive behavior during verbal deception. *International Journal of Behavioral Development, 26,* 436–444.

Talwar, V., & Lee, K. (2002b). Emergence of white lie-telling in children between 3 and 7 years of age. *Merrill-Palmer Quarterly, 48,* 160–181.

Talwar, V., Lee, K., Bala, N., & Lindsay, R. C. L. (2002). Children's conceptual knowledge of lie-telling and its relation to their actual behaviors: Implications for court competence examination. *Law and Human Behavior, 26,* 395–415.

Talwar, V., Lee, K., Bala, N., & Lindsay, R. C. L. (2004). Children's lie-telling to conceal a parent's transgression: Legal implications. *Law and Human Behavior, 28*, 411–435.

Talwar, V., Murphy, S., & Lee, K. (2007). White lie-telling in children for politeness purposes. *International Journal of Behavioral Development, 31*, 1–11.

Welsh, M. C., & Pennington, B. F. (1988). Assessing frontal lobe functioning in children: Views from developmental psychology. *Developmental Neuropsychology, 4*, 131–149.

Welsh, M. C., Pennington, B. F., & Groisser, D. B. (1991). A normative-developmental study of executive function: A window on prefrontal function in children. *Developmental Neuropsychology, 7*, 131–149.

Wilson, A. E., Smith, M. D., & Ross, H. D. (2003). The nature and effects of young children's lies. *Social Development, 12*, 21–45.

Wimmer, H., Gruber, S., & Perner, H. (1984). Young children's conception of lying: Lexical realism-moral subjectivism. *Journal of Experimental Child Psychology, 37*, 1–30.

Wimmer, H., & Hartl, M. (1991). Against the cartesian view on mind: Young children's difficulty with own false beliefs. *British Journal of Developmental Psychology. Special issue: Perspectives on the child's theory of mind, 9*, 125–138.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs. Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*, 103–128.

Zelazo, P. D., Carter, A., Reznick, J. S., & Frye, D. (1997). Early development of executive function: A problem-solving framework. *Review of General Psychology, 1*, 1–29.

Zelazo, P. D., & Muller, U. (2002). Executive function in typical and atypical development. In U. Goswami (Ed.), *Blackwell handbook of childhood cognitive development* (pp. 445–470). Malden, MA: Blackwell.

Child Development, July/August 2008, Volume 79, Number 4, Pages 882–898

Moral Development in a Violent Society: Colombian Children's Judgments in the Context of Survival and Revenge

Roberto Posada and Cecilia Wainryb
University of Utah

Ninety-six Colombian children (mean age = 7.7 years) and adolescents (mean age = 14.6 years) made judgments about stealing and physical harm in the abstract and in the context of survival and revenge. All participants judged it wrong to steal or hurt others because of considerations with justice and welfare, and most also judged it wrong to engage in such actions even when they can aid in survival. Their judgments in the context of revenge were more mixed, with a sizable proportion endorsing stealing and hurting in that condition. Furthermore, the majority expected that people would steal and hurt others in most situations. Significant age differences were also found. The consequences of political violence for moral development are discussed.

As reported by the UN Office of the Special Representative of the Secretary General for Children and Armed Conflict, children in approximately 50 countries around the world currently grow up in the midst of armed conflict and its aftermath. In the past decade, 6 million children were injured due to war and 2 million were killed. Millions of children have been uprooted as a consequence of armed conflict and human rights violations, seeking safety abroad or within their own borders; indeed, well over half of the world's displaced people are children. Currently, it is estimated that there are 10 million children refugees worldwide; an additional 13 million children are internally displaced within their own countries (United Nations, 2006). In Colombia alone, the site of the present study, almost 2 million children have been forcibly displaced from their homes and towns during the past 15 years. Appalling and shocking in and of themselves, these figures also carry the disturbing implication that more and more of the world's children are being sucked into a bleak moral vacuum—a psychological space devoid of basic human rights and values. How might children's development be altered by the violence, lawlessness, and deprivation to which they are exposed?

Psychologists have long been concerned with the effects that long-term exposure to violent conditions has on children's psychological well-being. While much of this research has been conducted in violent communities within the United States (e.g., Buka, Stichick, Birdthistle, & Earls, 2001; Richters & Martinez, 1993), studies have also been conducted in countries at war, such as Israel and the occupied West Bank and Gaza, Lebanon, and South Africa (e.g., Barber, Schluterman, Denny, & McCouch, 2006; Macksoud & Aber, 1996; Punamäki, 1996; Straker, Mendelson, Moosa, & Tudin, 1996). Most research efforts in this regard have used a trauma model and measured the consequences of chronic exposure to violence in terms of mental health outcomes. Findings concerning children's specific patterns of symptoms have varied across studies as a function of the domains of adjustment targeted and the characteristics of the samples studied. In general, however, children's reactions associated with exposure to chronic violence have included anxiety, depression, psychosomatic disturbances, and other symptoms that have come to be associated with posttraumatic stress (Richters & Martinez, 1993).

Still, children may respond in complex ways that are not necessarily captured by the posttraumatic stress disorder (PTSD) diagnostic criteria, and the exclusive focus on psychiatric symptoms might be mistakenly taken to imply that children who do not receive a diagnosis are unaffected. Nonetheless, in comparison to the large volume of work documenting the clinical consequences of exposure to political violence, very little is known about the nonclinical impact that growing up in such conditions may have on children. Indeed, in their commentary to the Special Issue devoted to children growing up in the midst of political violence that appeared a decade ago in *Child Development*, Cairns

---

This article is based on the first author's master's thesis. The research was supported by two fellowships (2003–2004 and 2004–2005) from the Barbara and Norman Tanner Center for the Prevention of Violence.

Correspondence concerning this article should be addressed to Cecilia Wainryb, Department of Psychology, University of Utah, 380 South 1530 East, Salt Lake City, UT 84112. Electronic mail may be sent to wainryb@psych.utah.edu.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7904-0006

and Dawes (1996) noted the scarcity of research on the significance of war-related trauma on normative development and urged researchers to move beyond the documentation of distress symptoms and place a greater emphasis on examining the impact that war, violence, and dislocation has on children's development (see also Richters & Martinez, 1993).

Even as a number of researchers have alluded to the potential effects that war and political violence are likely to have on children's moral development (e.g., Cairns & Dawes, 1996; Dawes, 1994; Garbarino & Kostelny, 1996; Leavitt & Fox, 1993; Macksoud & Aber, 1996; Punamäki, 1996), systematic research about these issues has been limited and its findings have been mixed, with some studies indicating that the moral development of children living in the midst of political violence is truncated, as evidenced in lower levels of moral reasoning, and others suggesting that the moral development of these youths is not negatively affected (for comprehensive and critical reviews of this research, see Cairns, 1996; Garbarino & Kostelny, 1993). The mixed nature of these findings may in part be due to the fact that the existing evidence was obtained largely on the basis of global assessments of stages of moral reasoning, with the underlying question being whether children in communities characterized by political violence reason at lower, less mature, stages than children in nonviolent communities.

The limitations of global assessments of moral stages have been exposed by moral development research conducted in the past several decades (Turiel, 1998). Two reliable findings emerging from this research are potentially relevant to the present discussion. First, it has been shown that rather than developing principled moral concepts out of less mature nonmoral concerns (such as a focus on reward and punishment), children construct moral concepts on the basis of their social interactions, and thus even young children are likely to judge that it is wrong to, for example, hurt others or steal not due to fear of punishment but because of a concern with fairness and the welfare of others. The other, complementary finding was that, along with their moral understandings, children also construct understandings about other aspects of social interactions and, when facing concrete social situations, vary in how they weigh conflicting considerations (see reviews by Smetana, 2006; Turiel, 1998).

These issues are of particular relevance to the study of moral development among war-affected children because their social interactions may not facilitate the development of moral concepts. Rather, war-affected children, exposed to conditions of violence, poverty, lawlessness, and displacement, are likely to face situations in which their moral principles about, for example, not hurting others or not stealing come into potential conflict with other significant considerations, such as their own needs for food and shelter or their wish to secure justice or retribution for having been wronged. Thus, questions about the overall stage of moral reasoning of children exposed to political violence, or global comparisons between children exposed and not exposed to violence in terms of their moral stages, may not be sufficient for identifying the unique features and vulnerabilities in the moral thinking of war-affected children. Rather, two main questions are of interest. The first, most basic, question is whether, in spite of the violence and lawlessness to which they are exposed, war-affected children develop prescriptive and generalizable moral concepts bearing on justice and welfare. If they do, the second question is how they might bring these moral concepts to bear on multifaceted situations—situations that feature considerations that are relevant to life in the midst of political conflict and that might represent compelling reasons for breaching moral principles, such as concerns with survival or revenge.

Our hypotheses concerning the first question, of whether war-affected children develop moral concepts, rely on evidence about the development of moral concepts in the context of sociocultural conditions that may also be seen as adverse. Research has shown that children growing up in traditional hierarchically organized societies develop concepts about rights and autonomy in spite of the fact that their culture's discourse and practices tend to discourage such conceptions (Turiel & Wainryb, 1998), and girls and women in patriarchal societies, who are subjected to practices that enforce gender inequalities, nevertheless make judgments about the unfairness of those practices (Wainryb & Turiel, 1994). Similarly, despite explicit socialization into Communist ideology, adolescents in Mainland China develop universalizable notions of accountability and impartiality and make judgments favoring democratic principles (Helwig, Arnold, Tan, & Boyd, 2003, 2007). Together, these findings suggest that the development of moral concepts is not merely determined by cultural and societal conditions but grounded in the intrinsic features of harmful and unjust actions (Turiel, 1998). Although cultural ideologies are not the same as actual exposure to violence and injustice, our expectation was that children exposed to political violence would nevertheless develop prescriptive and generalizable conceptions of welfare and justice.

To more fully understand the second question, of how war-affected children bring their moral concepts to bear on situations underscoring concerns with survival or revenge, it might be useful to examine not only their *moral* judgments (such as whether they think it right or wrong to, for example, steal in order to fulfill basic material needs) but also their *psychological* judgments (such as their expectations about how people are likely to act or feel when facing dire economic need). Recent research has indeed demonstrated that assessments of both psychological and moral judgments yield a more complete picture of children's actual moral experiences (Wainryb, 2004; Wainryb & Brehl, 2006; Wainryb, Brehl, & Matwin, 2005).

A number of psychological judgments are of interest. These children's expectations about what people are *likely* to do (as distinct from their judgments about what people *should* do) are interesting because war-affected youths function in the midst of chronic violence and lawlessness. Therefore, their moral experiences are likely to encompass a tension between their prescriptive judgments about whether, for example, stealing is right or wrong in a particular context and their knowledge and expectations about whether people (or even themselves) are likely to steal in that context. Although research in this regard is nonexistent, our hypothesis was that war-affected children would expect people to steal and perhaps even to engage in physical harm more often in situations of survival and revenge than in situations in which those concerns are not present. Children's expectations concerning the emotional consequences of engaging in behaviors such as stealing or hurting in specific contexts are also potentially interesting because affective reactions are a salient feature of children's experiences of moral transgressions (Arsenio, Gold, & Adams, 2006). Research has shown that children's recognition of victims' sadness or anger, deemed central to their moral thinking, emerges very early on. Their understanding that transgressors may experience shame or guilt emerges relatively late, and up to about the age of 5 or 6, children tend to think that transgressors feel happy; when present in late childhood or adolescence, however, such "happy victimizer" attributions are predictive of restricted social competence or even psychopathology (Arsenio et al., 2006). Given the importance of these psychological understandings for moral functioning, it bears asking how children who have experienced widespread violence and injustice might construe the emotional experiences of those who steal or inflict harm because of concerns with survival and revenge.

In regard to their moral judgments about stealing and harm in such contextualized situations, previous research (Helwig, 1995; Turiel & Wainryb, 1998) led us to expect that war-affected children would judge those acts to be morally wrong in unconflicted situations but would endorse stealing more frequently in situations underscoring matters of survival. It was harder to predict whether war-affected children would endorse the infliction of physical harm on others in situations underscoring concerns with revenge. While research conducted with typically developing children suggests that by middle childhood, children condemn physical retaliation and revenge (e.g., Smetana, Campione-Barr, & Yell, 2003), it has also been shown that, when provocation occurs, aggressive children and adolescents tend to view physical aggression as morally justified (Astor, 1994; Gibbs, 2003).

One additional question of interest concerns the potential age and gender differences in the judgments of war-affected children. Previous research suggests that the long-term effects of violence exposure are moderated by age and gender. Given that the maturity of physiological systems and cognitive and sociocognitive abilities may pose both vulnerabilities and protective features at each level of development, it is not surprising that the data on age differences in children's responses to violence are inconsistent, with some studies suggesting that younger children show more disturbances and others finding that adolescents exhibit the most symptoms (e.g., Wright, Master, & Hubbard, 1998). Thus, although moral development research has shown that children develop moral concepts at a young age, it is hard to predict what differences, if any, may exist in the reasoning of children and adolescents growing up in a violent society. The potential effects of gender are similarly complicated, as research suggests that boys and girls respond differently to trauma, with externalizing problems being more prevalent among boys and internalizing problems among girls (Wright et al., 1998).

Method

Setting and Participants

Colombia has been in a state of civil war for more than 50 years, with leftist guerrilla groups and rightist paramilitary groups waging war against the government and against each other. The ongoing conflict has been characterized by widespread violence, resulting in one of the highest kidnapping and homicide rates worldwide (World Bank, 2000). As a result of this conflict, nearly 4 million people (out of a population of 44 million) have been forced to leave their homes and towns, seeking refuge in the big cities. More than 400,000 of these internal refugees have relocated in shanty towns in the outskirts of Bogotá—a startling number given that the capital's total population is 8 million—creating peripheral rings of squatter settlements characterized by high population density, poor housing, lack of public services, and inadequate nutrition. Employment opportunities for internally displaced persons are particularly low, largely because more than 80% of them had been peasants prior to their displacement and lack knowledge about urban job opportunities. War-displaced children also face significant hurdles in continuing their education, both as a direct consequence of their status as displaced, as when they are required to produce forms of identification they no longer possess, and because they cannot meet the expenses associated with attending school. Thus, it is not surprising that only about 8% of displaced children are enrolled in school as compared to 93% of school-age children native to those communities (Human Rights Watch, 2006).

The present study was carried out in Usme, one of southern Bogotá's poorest slums (population 230,000) featuring one of the largest concentrations of displaced persons. Usme is a vast warren of concrete homes and plywood-and-aluminum shacks. Only some roads are paved, and in many places, people have to walk through mud and sewage. Sanitation is poor and access to water and electricity sporadic (e.g., during the time we conducted interviews, the water supply was cut off for days at a time). By Colombian socioeconomic status (SES) standards, individuals residing in Usme are classified in the "low" (*nivel bajo*, Strata 2) and "low-low" (*nivel bajo-bajo*, Strata 1) levels, characterized as living under extreme poverty. Usme is considered one of the most dangerous localities in Bogotá, with a rate of 3.3 violent deaths per 10,000 (Red Bogotá, 2006). Participants in the study were 96 displaced children (mean age = 7.7, range = 6.0–9.8), and adolescents (mean age = 14.6, range = 13.0–16.9), evenly divided by gender and age. About 40% of participants reported having lived in Usme for 1 year or longer, and the remaining had been in Usme for less than a year; however, because displaced persons often move from one neighborhood to another, the total time since *first* displacement is not known. Although most participants reported being enrolled in school (89%) and not working (88%), these figures may not be reliable because children are expected to attend school, and it is illegal for them to hold a job. Although reliable figures could not be obtained, many of the participants in the study lived alone, with siblings or peers, had no parents, or lacked contact with them.

Procedures and Assessments

Displaced children and adolescents residing in or near Usme were invited to participate in the study by the director of Cooperemos D.P.S., a nongovernmental organization (NGO) that assists displaced persons in Bogotá. Children interested in participating were asked to drop by the community center at prearranged times. Children were not offered any form of payment for their participation, but food was made available to them before and after the interview (food was not restricted to those who participated in the interview).

Waiver of parental consent was sought, as most of the youngsters likely to participate in the study either had lost their parents during the violence preceding their displacement or were solely responsible for their own survival. This is not uncommon in developing countries in South America, where children as young as 5–6 can be seen in public places, engaging in a variety of survival activities (selling candy or newspapers, washing cars, shining shoes, entertaining passers-by, begging, and scavenging; Rafaelli & Larson, 1999). In lieu of parental consent, the NGO's director agreed to act as the children's advocate. As such, the director granted consent on behalf of the entire group of children; this procedure was deemed most desirable as it did not require generating a record of the individual children's names. In addition, the agency's director (or a person designated by him) witnessed each child's assent procedure. Verbal, rather than written, consent was obtained from participants, as many of these youngsters are unable to read and write at the level required to fully understand even a simplified written document. In addition, given the ubiquity of illegal activity (including by children), youngsters are reluctant to give their full names (typically, the only agencies that require full names are government institutions where youth may be held against their will).

Following the assent procedure, children were interviewed in Spanish by the first author. Participants were given a two-part interview modeled after previous related research (e.g., Helwig, 1995; Turiel & Wainryb, 1998); the interview was conducted in one session of approximately 45 min. In the first part (general assessments), they were asked a series of questions tapping their general conceptions concerning stealing and inflicting physical harm. The purpose of these assessments was to determine whether children judge instances of harm and injustice as moral concerns, universally applicable across social contexts and not contingent on existing laws. In the second part (contextualized assessments), participants were asked to make psychological and moral judgments about stealing and inflicting physical harm in specific contexts featuring concerns with survival and revenge. The purpose of these assessments was to determine whether, and how, general moral concepts are applied in situations involving conflicts with considerations that may serve, in the sample studied, as reasons for breaching moral principles. Interviews were tape-recorded and later transcribed for analysis. Following the interview, a measure of participants' exposure to violence was obtained. Details about each set of assessments follow.

General assessments. Participants were asked to make judgments about the acceptability of stealing and inflicting physical harm ("Is it okay or not okay to [take other people's things/shove, kick, or hurt others]? Why?"). Next, they were asked whether they endorse laws against stealing and inflicting harm in their own country ("Do you think that there should be a rule against [taking other people's things/shoving, kicking, or hurting others]?") and in other countries ("Do you think that there should be a rule like that in all countries?"), and whether their judgments about stealing and physical harm are contingent on laws or common practice ("If there were no rules against [taking . . ./shoving . . .], would it be okay or not okay for someone to [take . . . /shove . . .]?" "In a country where it is very common for people to [take.../shove...], do you think it would be okay or not okay for someone to do that?"). All assessments were given first for stealing and next for inflicting harm.

Contextualized assessments. Children's psychological and moral judgments about stealing and inflicting physical harm were subsequently assessed in two contextualized conditions in which concerns with either survival or revenge were underscored; a baseline (unconflicted) condition, in which moral concerns were not pitted against other considerations, was also included for comparison. Examples of stimuli follow:

Baseline condition: Luis is 14 years old; he lives in northern Bogotá. One evening he was walking back home from the movies; the street was deserted. As he started crossing the street, he saw a guy about his age who was walking a bike. It was a good bike and it looked almost new; Luis thought that was a cool bike. His parents had told him that he could get a bike like that for his next birthday but that was still 8 months away, and Luis had wanted a bike for a long time.

Survival condition: Juan is 13 years old; he lives in Soacha. One evening he was walking back home; it was getting dark, and the street was deserted. Juan was feeling worried. A friend helped him get a job delivering newspapers and Juan needs the job to help feed his younger brothers and sisters, but to deliver newspapers he needs a bike. Bikes are awfully expensive and Juan does not have enough money to buy one. As he's thinking about this, he sees a guy about his age across the street; the guy is walking a bike. It's a good bike and it looks almost new.

Revenge condition: Julio is 15 years old; he lives in Ciudad Bolivar. One evening he was walking back home from the store. There were no people in the street. Julio was thinking about his old friends and about the house where he used to live in with his family before they had to move. As he looks up to cross the street, he sees a guy not much older than him. Julio had seen that guy before, he's one of "them," the people who hurt his father and his brother and forced his family to move. The guy is walking a bike; it's a good bike, and it looks almost new.

For each condition, psychological and moral judgments were assessed. Participants were first asked to predict the protagonist's likely course of action (e.g., "What do you think Juan is going to do? Why do you think he might do that?"). Next, they were told that the protagonist engaged in stealing (e.g., "Juan followed him quietly and when the guy was distracted, Juan grabbed the bike and rode away") and were asked to predict what the protagonist and victim might have felt (e.g., "How do you think Juan felt after he grabbed the other's guy bike and rode away?" "How do you think the other guy felt when Juan grabbed his bike?") and to judge the protagonist's behavior (e.g., "Do you think it was okay or not okay for Juan to grab the guy's bike? Why?"). Finally, participants were told that the protagonist also inflicted physical harm on the other character (e.g., "Let's say that Juan followed the guy and when he had an opportunity, he shoved him down to the ground, kicked him really hard, grabbed the bike and rode away."), and the same assessments (protagonist's emotion, victim's emotion, and protagonist's behavior) were obtained.

The three conditions were presented in a fixed order (baseline, survival, and revenge). To allow for generalizability across content areas, and to sustain participants' interest, three comparable versions were designed. In addition to the scenario dealing with a bike, another scenario dealt with a jacket and a third one with a boom box. The "jacket" and "boom box" versions of the baseline and revenge conditions were identical to the "bike" version (i.e., in the baseline condition, protagonist desired the jacket/boom box; in the revenge condition, protagonist noticed a member of the group responsible for her/his displacement wearing a jacket/carrying a boom box). For the purpose of presenting a jacket or boom box as related to survival, the "jacket" version stated that: "Juan knows that he has grown up a lot in the last six months and his clothes don't fit him anymore, they're too tight, especially his jacket. But he can't buy a new jacket now; at home they have barely enough money for food." The "boom box" version of the survival condition stated that: "Juan's grandma is so sick, she's stuck at home all alone all day. And now that the old radio broke down, there's nothing to keep her company. But he can't buy a new radio for her; at home they have barely enough money for food." Each participant heard each condition with a different scenario (e.g., baseline/bike; survival/jacket; revenge/boom box). Half of the participants in each age and gender group heard the stimuli with two male characters and the other half with two female characters. The interview was written in English, and then translated into Spanish by the first author, and back-translated by a bilingual research assistant.

Exposure to violence assessment. For the purpose of documenting the type and extent of violence exposure in this sample, participants were given the Spanish version of the Exposure to Violence interview (ETV; Selner-O'Hagan, Kindlond, Buka, Raudenbush, & Earls, 1998). The ETV is a structured interview assessing 24 types of violence (e.g., being attacked with a weapon, hearing gunfire) experienced as witness and as victim; six items (e.g., bearing on natural catastrophes, suicide) were omitted because they were deemed irrelevant. Exposure scores as "witness" and "victim" and a total ETV score were calculated based on the number of items endorsed. The ETV has a reported test – retest reliability of .75 – .94 and internal consistency indices, as measured by Cronbach's alphas, from .68 to .93.

Scoring and Reliability

General assessments. Two aspects of participants' judgments were scored using categories adapted from the widely used Davidson, Turiel, and Black (1983) scoring system. Evaluations were scored on a 3-point scale. *Evaluations* that stealing and inflicting harm (a) are wrong, (b) should be illegal in their own country, (c) and in other countries, (d) regardless of existing rules, (e) and of common practice, were assigned a score of 3; evaluations that stealing and inflicting harm are not wrong, should not be illegal in their own or other countries, or that their status is contingent on existing rules or common practice were assigned a score of 1; mixed evaluations (e.g., it is both right and wrong) on any of these assessments were assigned a score of 2. *Justifications* were scored using the following categories: welfare (e.g., "because it's wrong to punch people, it can hurt them badly"), justice (e.g., "because it isn't fair to just take something that doesn't belong to you"), and rules/punishment (e.g., "because the police will throw you in jail"). Multiple justifications were scored in terms of the proportional use of each category.

Contextualized assessments. Categories used for scoring the *expected course of action* were derived from scoring pilot data and 20% of this study's protocols. Responses were first scored as non violence (e.g., "she will borrow it from someone"), stealing (e.g., "he will just take it and run"), or violence against persons (e.g., "he will knock him down and hit him"). Next, *reasons* given for the expected course of action were scored using categories such as needs (e.g., "because his brothers and sisters were hungry"), retribution (e.g., "because this guy hurt his family and took away everything, so he's getting his revenge"), hate (e.g., "because she hates these people"), justice (e.g., "because it wouldn't be fair if he just stole it"), likes and desires (e.g., "because she really liked that bike"), and prudence (e.g., "because if he does something to him, then the other guy's friends will come and hurt him worse"). Multiple reasons were scored in terms of the proportional use of each category. The protagonist's and victim's *emotions* were scored using categories adapted from previous scoring systems (Arsenio et al., 2006), including guilt/shame, fear, sadness, anger, hate, and happiness. Responses combining two negative emotions (e.g., "she felt guilty and also angry") were scored in terms of the proportional use of each emotion category. Responses combining a positive emotion and a negative emotion (e.g., "he felt happy but also must've felt guilty") were scored into one of the two mixed-emotion categories, suggesting a combination of satisfaction and a moral emotion (happiness + guilt/shame) or of satisfaction and a nonmoral emotion (happiness + fear or hate). The evaluations of the *protagonist's behavior* were scored on a 3-point scale as *positive* (1), *mixed* (2), or *negative* (3). Justifications were scored using categories adapted from the Davidson et al. (1983) system, including welfare/justice (e.g., "it's not fair what Ines did, it's not her boom-box, it's not fair to just steal it"), rules/punishment (e.g., "it's not okay because the other guy could go to the police and he might get caught"), and retaliation (e.g., "I think it's okay, he deserves it, because he was the one that hurt Julio's family in the first place"). Multiple justifications were scored in terms of the proportional use of each category.

Scoring reliability was assessed through the independent rescoring of 20% of the protocols by a second Spanish-speaking scorer. Interjudge agreement was 99% (Cohen's $\kappa = .95$) for general assessments, 99% (Cohen's $\kappa = .98$) for expected course of action, 86% (Cohen's $\kappa = .82$) for reasons, 93% (Cohen's $\kappa = .91$) for emotions, 99% (Cohen's $\kappa = .95$) for moral evaluations, and 88% (Cohen's $\kappa = .86$) for justifications.

Results

Exposure to Violence

Findings from the assessment of violence exposure are presented first so as to better frame the main findings of the study. Table 1 shows the percentage of participants reporting having been witnesses to or victims of different types of violent events. ETV scores ranged from 1 through 17, with a mean of 7.30 (*SD* = 3.3) and a mode of 9.00. As shown in Table 1, nearly all participants (92%) reported having heard gunfire. A majority also reported witnessing someone being hit or punched (83%), chased (63%), or threatened with bodily harm (61%), and hearing that someone was killed (65%). No less important are the findings about participants who reported having witnessed the most severe types of violence, such as finding a dead body (48%), or witnessing someone being killed (30%), attacked with a weapon (39%), or being shot at (32%) or shot (34%). Indeed, the large majority (88%) of participants reported being exposed to at least one of the most severe types of violence, 70% reported being exposed to two different types of severe violence, and 52% to three or more. As would be expected, victimization rates were lower than witnessing rates. About half of the participants reported having been hit or punched (46%), and a minority also reported having been chased (29%), threatened (28%), or attacked with a weapon (8%). Analyses of variance (ANOVAs) of the mean ETV scores (also reported in Table 1) by age and gender yielded significant age effects, with adolescents reporting higher exposure rates in general, $F(1, 92) = 11.79$, $p < .001$, $\eta_p^2 = .12$, as well as higher rates for witnessing, $F(1, 92) = 11.18$, $p < .001$, $\eta_p^2 = .11$, and victimization, $F(1, 92) = 4.70$, $p < .05$, $\eta_p^2 = .05$. A gender effect was found only for victimization, with males ($M = 1.44$, $SD = 1.3$, range $= 0-5$) reporting having been victims of more types of violent incidents than females ($M = 0.85$, $SD = 1.0$, range $0-3$), $F(1, 92) = 6.40$, $p < .01$, $\eta_p^2 = .07$.

Table 1
Percentage of Participants Reporting Having Witnessed and Having Been Victims of Different Types of Violent Acts and Mean Scores of Exposure to Violence

Type of violence	Witness	Victim
Was hit/slapped/punched	83	46
Was chased	63	29
Was threatened	61	28
Was attacked with a weapon	39	8
Was shot	34	1
Was shot at	32	4
Found a dead body	48	—
Saw someone being killed	30	—
Heard that someone was killed	65	—
Heard that someone was shot	29	—
Heard gunfire nearby	92	—
Witness ETV score, *M* (*SD*)		
Children	4.94 (2.2)	
Adolescents	6.50 (2.4)	
Victim ETV score, *M* (*SD*)		
Children	0.90 (0.9)	
Adolescents	1.40 (1.4)	
Total ETV score, *M* (*SD*)		
Children	6.2 (2.7)	
Adolescents	8.4 (3.6)	

Note. The dash indicates the areas do not exist in the Exposure to Violence interview (ETV).

Analytic Strategy of Interview Data

ANOVAs were used to analyze participants' evaluations of the general assessments and of the protagonist's behavior, all scored in terms of 3-point ordinal scales. GLM procedures were adopted for analyzing participants' justifications, expected course of action, reasons, and emotions (each scored in terms of the proportional use of various categories). Although loglinear- and structural equation modeling (SEM)-based procedures are typically deemed most appropriate for analyzing categorical data, the sparseness of this study's data set poses unique challenges to these procedures because they rely on computations within cells, thus crashing when encountering zeros. Analyses made using loglinear procedures come to a halt due to the impossibility of logging zeros (see, e.g., Wainryb, Shaw, Laupa, & Smith, 2001). Sparseness becomes a challenge for SEM when modeling the data by specifying relevant latent factors. Indeed, when attempts were made to model the present data, the latent factors resulted in untrustworthy solutions (e.g., negative variance) and no model converged with acceptable results. Although analyses using a one-parameter RASCH model worked in some cases (yielding findings identical to those obtained with GLM), they too failed when encountering cells with zeros. GLM gets around this problem because it computes variability across all cells, thereby capitalizing on the fact that sparseness is cell dependent. (Note that sparseness in this case is theoretically predicted and is not a result of a coding scheme that is inappropriate or failing to capture the data.) Although, in relying on GLM, assumptions about the nature of the covariance matrix are violated, evidence indicates that this violation of assumptions is likely to produce a more conservative test of the hypotheses (Kupek, 2006). Furthermore, it should be noted that where more appropriate tests (such as RASCH analyses) can be used, the results are identical with those produced by GLM.

Preliminary analyses by "version" yielded no significant effects or interactions; "version" was thus dropped from all subsequent analyses. For all analyses, post hoc comparisons using Bonferroni *t* tests were performed to test for significant within-subjects effects. Checks for skewness and kurtosis were conducted and, where appropriate, arcsine and logarithmic transformations were used. Analyses with transformed and untransformed data yielded identical results; results are presented with the untransformed data.

General Assessments

All participants judged that it is wrong to steal and hurt others. Repeated measures ANOVAs by age, gender, and transgression (stealing – inflicting harm), with transgression as a repeated measure, were performed on participants' evaluations of the other moral criteria. As shown in Table 2, the large majority of participants of all ages stated that there should be laws against stealing and causing physical harm to others and that it would be wrong for people to steal or hurt others even if such laws did not exist or if those actions were common in a specific context. Although the majority also judged that laws against stealing and hurting should exist in all countries, adolescents (96%) did so more often than children (75%), $F(1, 92) = 9.06$, $p < .01$, $\eta_p^2 = .09$. No significant effects or interactions involving gender were found. In justifying their evaluations, the large majority of participants referred to moral concerns with justice and welfare. A multivariate analysis of variance (MANOVA) conducted on the proportional use of each justification category by age, gender, and transgression yielded a significant effect for type of transgression ($p < .001$). Follow-up ANOVAs indicated that concerns with fairness were raised to justify the negative evaluations of stealing (62%) more often than those of inflicting harm (22%), $F(1, 91) = 42.53$, $p < .001$, $\eta_p^2 = .32$, whereas concerns with welfare justified negative evaluations of harm (55%) more often than those of stealing (13%), $F(1, 91) = 60.32$, $p < .001$, $\eta_p^2 = .40$. Only a minority of participants justified their negative evaluations of stealing (24%) and harming (20%) on the basis of nonmoral reasons (i.e., rules and punishment).

Table 2

Evaluations of Moral Criteria for Stealing and Inflicting Harm (Percentages and Means)

Moral criterion	Stealing	Inflicting harm
It is wrong		
%	100	100
M (SD)	3.00 (0.00)	3.00 (0.00)
It should be illegal in our country		
%	99	100
M (SD)	2.98 (0.20)	3.00 (0.00)
It should be illegal in all countries		
%	86	97
M (SD)	2.71 (0.71)	2.94 (0.35)
It would be wrong even if legal		
%	100	100
M (SD)	3.00 (0.00)	3.00 (0.00)
It would be wrong even if commonly practiced		
%	99	100
M (SD)	2.98 (0.20)	3.00 (0.00)

Note. Means are based on a 3-point scale; higher numbers indicate decreasing acceptability for acts and increasing acceptability for laws.

Contextualized Assessments

Expected course of action. Participants were asked to predict what the protagonist might do; their responses were coded as behaviors involving no violence, stealing, or violence against persons (see Table 3). A MANOVA conducted on the proportional use of each course of action by age, gender, and condition (baseline – survival – revenge), with condition as a repeated measure, yielded significant effects for age ($p < .05$) and condition ($p < .001$), as well as an Age × Condition interaction ($p < .001$). No significant effects or interactions involving gender were found. In general, participants expected a nonviolent course of action (e.g., "she will ask her if she can borrow the bike") more frequently in the baseline condition (54%) than in the survival (13%) or revenge (16%) conditions, $F(2, 182) = 38.93$, $p < .001$, $\eta_p^2 = .30$, and expected stealing (e.g., "he will probably snatch the boom-box and ran") more frequently in the survival condition (85%) than in the baseline (45%) or revenge (57%) conditions, $F(2, 182) = 24.87$, $p < .001$, $\eta_p^2 = .22$. Notably, however, nearly half of the participants

Table 3

Expected Course of Action, by Condition and Age (Percentages)

Course of action	Baseline Children	Baseline Adolescents	Survival Children	Survival Adolescents	Revenge Children	Revenge Adolescents
Nonviolence	69	40	15	10	17	15
Stealing	31	59	82	89	66	48
Violence against persons	0	1	3	1	18	38

Note. Due to rounding, percentages may not add up to 100.

(45%) expected the protagonist to steal even in the baseline condition (i.e., when no concerns with survival or revenge were underscored)—an unexpected finding that was more marked among adolescents. Indeed, as indicated by significant Age × Condition interactions, in the baseline condition, adolescents expected stealing more often than children, $F(2, 182) = 7.31$, $p < .001$, $\eta_p^2 = .07$, whereas children expected a nonviolent course of action more often than adolescents, $F(2, 182) = 3.88$, $p < .05$, $\eta_p^2 = .04$. Although infrequent overall (10%), violent courses of action (e.g., "he'll be so angry he'll probably kick the guy"; "I bet he will want to hurt him and see him suffer") made up nearly one third (28%) of responses in the revenge condition but only 1%–2% of responses in the other two conditions, $F(2, 182) = 46.14$, $p < .001$, $\eta_p^2 = .34$. This was, again, more marked among adolescents (38%) than among children (18%), $F(2, 182) = 6.80$, $p < .001$, $\eta_p^2 = .07$.

Participants were also asked to explain why they expected such a course of action (see Table 4). A MANOVA conducted on the proportional use of each reason by condition, age, and gender yielded significant effects for age ($p < .01$) and condition ($p < .001$), as well as an Age × Condition interaction ($p < .001$). As shown in Table 4, the expected course of action in the survival condition was justified largely (76%) in terms of needs (e.g., "I think it's because she doesn't have any money and she's desperate to help her family") and the expected course of action in the revenge condition was justified in terms of retribution (e.g., "because that guy hurt his family so he's getting back at him"; 56%) and, to a lesser extent, hate (e.g., "I think because you get to hate people if they take everything away that you love"; 12%). Reasons related to need were more common in the survival condition than in the other conditions, $F(2, 182) = 160.12$, $p < .001$, $\eta_p^2 = .64$, and reasons related to both retribution, $F(2, 182) = 159.19$, $p < .001$, $\eta_p^2 = .64$, and hate, $F(2, 182) = 20.52$, $p < .001$, $\eta_p^2 = .18$, were more common in the revenge condition than in the other conditions. As indicated by Condition × Age interactions, adolescents, more often than children, referred to needs in the survival condition, $F(2, 182) = 6.05$, $p < .01$, $\eta_p^2 = .06$, and to retribution, $F(2, 182) = 3.10$, $p < .05$, $\eta_p^2 = .03$, and hate, $F(2, 182) = 9.34$, $p < .001$, $\eta_p^2 = .09$, in the revenge condition.

Table 4

Reasons for Expected Course of Action, by Condition and Age (Percentages)

Reasons	Baseline Children	Baseline Adolescents	Survival Children	Survival Adolescents	Revenge Children	Revenge Adolescents
Needs	13	3	67	84	12	3
Retribution	0	0	0	0	48	65
Hate	0	0	1	1	4	20
Justice	42	31	2	6	1	3
Likes/desires	31	59	26	7	22	1
Prudence	4	0	2	0	9	8
Other/unelaborated	10	6	2	2	4	0

Note. Due to rounding, percentages may not add up to 100.

Findings concerning the reasons for the course of action in the baseline condition were more unexpected. Whereas participants justified nonviolent expectations in terms of justice (e.g., "because it would be just wrong for her to grab it and run, it wouldn't be fair"), their expectations of stealing were justified, not in terms of needs, but in terms of likes and desires (e.g., "because he really wanted it and he didn't want to wait"). Justice reasons, $F(2, 182) = 35.95, p < .001, \eta_p^2 = .28$, and likes and desires, $F(2, 182) = 22.99, p < .001, \eta_p^2 = .20$, were given more frequently for the baseline condition than for the other conditions. As indicated by a significant Condition × Age interaction, $F(2, 182) = 12.75, p < .001, \eta_p^2 = .12$, however, adolescents provided reasons related to likes and desires more often than children in the baseline condition, but in the other two conditions children provided these reasons more frequently than adolescents.

Protagonist's emotions. Subsequent to eliciting their thoughts about the protagonist's expected course of action, participants were told that the protagonist (a) stole from or (b) incurred in physical violence against the other character and were asked to predict what the protagonist may have felt following each transgression. The distribution of emotions attributed to the protagonist is presented in Table 5. Our expectation was for participants to think that the protagonist might experience either negative moral emotions (e.g., guilt, shame) or a mixture of such moral emotions along with more positive emotions (especially in the survival and revenge conditions). As shown in Table 5, only about 40% in the baseline and survival conditions and less than 30% in the revenge condition expected the protagonist to feel negative moral emotions. Although participants often attributed mixed emotions to the protagonists, and did so in all conditions (36%–38%), about one third of the mixed emotions did not include a moral emotion but rather conveyed a mixture of happiness and either fear or hate.

The MANOVA conducted on the proportional use of each emotion category by condition, transgression, age, and gender (with condition and transgression as repeated measures) yielded significant effects for condition ($p < .001$) and age ($p < .001$), as well as a Condition × Age interaction ($p < .001$), and a significant effect for type of transgression ($p < .001$). Protagonists were said to feel guilt or shame, $F(2, 182) = 13.07, p < .001, \eta_p^2 = .13$, or a mixture of guilt or shame and happiness, $F(2, 184) = 5.77, p < .01, \eta_p^2 = .06$, more often in the baseline and survival conditions than in the revenge condition. By contrast, hate, $F(2, 182) = 16.03, p < .001, \eta_p^2 = .15$, or a mixture of hate or fear and happiness, $F(2, 184) = 9.38, p < .001, \eta_p^2 = .09$, were more commonly attributed to protagonists in the revenge than in the other two conditions. As indicated by significant Condition × Age interactions, adolescents attributed to protagonists in the revenge condition both hate, $F(2, 182) = 10.37, p < .001, \eta_p^2 = .10$, and happiness, $F(2, 182) = 6.43, p < .01, \eta_p^2 = .07$, more often than children did, whereas children, more often than adolescents, attributed to these protagonists guilt or shame, $F(2, 182) = 8.42, p < .001, \eta_p^2 = .09$. Both types of mixed emotions were more commonly attributed by adolescents than by children: Adolescents, more often than children, attributed to protagonists a mixture of happiness and guilt/shame in the baseline and survival conditions, $F(2, 184) = 5.16, p < .01, \eta_p^2 = .05$, and a mixture of happiness and hate/fear in the revenge condition, $F(2, 184) = 8.32, p < .001, \eta_p^2 = .08$.

Different emotions were also attributed to protagonists depending on the type of transgression they

Table 5
Protagonist's Emotion, by Condition and Age, and by Transgression Type (Percentages)

	Condition and age							
	Baseline		Survival		Revenge		Transgression type	
Emotion	Ch	Ad	Ch	Ad	Ch	Ad	Stealing	Harming
Guilt/shame	48	34	50	35	47	8	30	44
Hate	1	5	1	2	3	21	5	6
Fear	15	10	13	11	16	2	12	10
Happiness	10	4	13	3	9	17	12	7
Guilt/shame + happiness	19	42	18	33	18	18	32	17
Hate or fear + happiness	6	5	7	12	7	32	9	14
Other	2	1	0	4	0	2	1	2

Note. Due to rounding, percentages may not add up to 100. Ch = children; Ad = adolescents.

Table 6
Victim's Emotion, by Condition and Age (Percentages)

	Baseline		Survival		Revenge	
Emotion	Children	Adolescents	Children	Adolescents	Children	Adolescents
Sadness	80.8	46.6	80.6	50.9	85.4	42.2
Anger	6.3	36.0	5.3	30.8	6.8	41.6
Fear	6.8	15.2	10.0	17.2	6.8	13.7
Other	6.3	2.2	4.3	1.1	1.1	2.1

Note. Due to rounding, percentages may not add up to 100.

were said to have incurred. Whereas a mixture of happiness and guilt or shame was more commonly attributed to protagonists said to have stolen from, rather than inflicted harm on, the victim, $F(1, 92) = 18.74, p < .001, \eta_p^2 = .17$, a mixture of happiness and hate or fear was more commonly attributed to protagonist said to have inflicted harm on, rather than stolen from, the victim, $F(1, 92) = 4.12, p < .05, \eta_p^2 = .04$. Negative moral emotions (guilt and shame) were also more commonly attributed to protagonists who inflicted harm on, rather than to those who stole from, the victim, $F(1, 91) = 24.52, p < .001, \eta_p^2 = .21$.

Victim's emotions. Participants were also asked what the victim might have felt (see Table 6). The MANOVA involving the proportional use of each emotion category by condition, transgression, age, and gender yielded a significant effect for age ($p < .001$). Although sadness was the emotion most commonly attributed to victims across all conditions and transgression types, children referred to the victim's sadness more often than adolescents, $F(1, 87) = 38.54, p < .001, \eta_p^2 = .31$. Adolescents, in turn, attributed anger to victims more often than children, $F(1, 87) = 36.14, p < .001, \eta_p^2 = .29$.

Protagonist's behavior. Finally, participants were asked to evaluate the protagonist's behavior. An ANOVA by age, gender, condition, and transgression, with condition and transgression as repeated measures, was conducted on participants' evaluations. In general, participants negatively evaluated both stealing and inflicting harm across all conditions. However, as shown in Table 7, participants made less negative evaluations of the protagonist's behavior in the revenge condition than in the other two conditions, $F(2, 184) = 36.76, p < .001, \eta_p^2 = .29$. Although children and adolescents did not differ significantly in their evaluation of the protagonists' behaviors, there was a tendency for adolescents to be more accepting of these behaviors in the revenge condition, $F(2, 184) = 2.41, p < .09$. It is important to underscore that there were no participants who judged that it was okay for the protagonist to steal from or inflict harm on the other character in the baseline or survival conditions, and only 1 or 2 participants in each condition gave mixed evaluations (i.e., it is okay and not okay). For the revenge condition, by contrast, 19% approved of the protagonist's stealing and 21% approved of his/her inflicting physical harm, and an additional 3%–10% gave partially positive (mixed) evaluations of those behaviors.

Participants were also asked to justify their evaluations (see Table 8). The MANOVA involving the proportional use of each justification category by age, gender, condition, and transgression yielded significant effects for age ($p < .001$) and condition ($p < .001$). Participants' negative judgments of

Table 7
Mean Evaluations and Percentages of Negative, Mixed, and Positive Evaluations of the Protagonist's Behavior, by Condition and Transgression Type

	Condition and transgression type					
	Baseline		Survival		Revenge	
	Stealing	Harming	Stealing	Harming	Stealing	Harming
M (*SD*)	2.99 (0.10)	3.00 (0.00)	2.98 (0.14)	2.99 (0.10)	2.54 (0.79)	2.55 (0.82)
% negative evaluations	99	100	98	99	71	76
% mixed evaluations	1	0	2	1	10	3
% positive evaluations	0	0	0	0	19	21

Notes. Means are based on 1 = *okay*, 2 = *okay and not okay*, 3 = *not okay*.

Table 8
Justifications for Evaluations of the Protagonist's Behavior, by Condition (Percentages)

	Condition		
Justification	Baseline	Survival	Revenge
Welfare/justice	65	64	27
Rules/punishment	23	22	39
Retribution	0	1	24
Unelaborated	12	14	11

Note. Due to rounding, percentages may not add up to 100.

stealing and causing harm in the baseline and survival conditions were justified largely in moral terms. Indeed, moral concerns with welfare and justice were more common in the baseline and survival conditions than in the revenge condition, $F(2, 182) = 49.43$, $p < .001$, $\eta_p^2 = .35$. Still, it should be noted that, somewhat unexpectedly, around 20% of participants who made negative evaluations of the protagonist's behavior in the baseline and survival conditions, did so for nonmoral reasons (e.g., "it wasn't okay for him to steal the bike because he might get caught and get in trouble"). This unexpected tendency was even more marked in participants' reasoning about the revenge condition. In this case, even though between 71% and 76% of participants had evaluated the protagonist's stealing and inflicting harm negatively, only 27% did so for moral reasons (e.g., concerns with the victim's welfare or with justice); 39% relied, instead, on references to rules and punishment (e.g., fearing punishment or further retribution). Indeed, concerns with rules and punishment were more common in the revenge condition than in the other two conditions, $F(2, 182) = 11.02$, $p < .001$, $\eta_p^2 = .11$. Concerns with retribution were also more common in the revenge condition than in the other two conditions, $F(2, 182) = 37.99$, $p < .001$, $\eta_p^2 = .29$, and were used largely to justify participants' positive evaluations of the protagonist's behavior. Finally, significant age effects indicated that, in general, adolescents (62%) referred to concerns with welfare and fairness more often than children (42%), $F(1, 91) = 13.45$, $p < .001$, $\eta_p^2 = .13$, and children (37%) referred to concerns with rules and prudence more often than adolescents (19%), $F(1, 91) = 10.26$, $p < .01$, $\eta_p^2 = .10$.

Discussion

Undeniably, children and adolescents in this sample have been exposed to considerable violence. Although the rates of victimization were relatively low, all participants had witnessed some form of violence perpetrated on others, with an average of 6.2 and 8.4 different types of violence, respectively, for children and adolescents, and the majority (88%) had witnessed at least one of the most severe and disturbing events, such as finding a dead body or seeing someone being shot. The finding that the prevalence rates of victimization, as compared to witnessing, were low is consistent with findings about children's exposure to violence in the United States (e.g., Richters & Martinez, 1993)—findings that have also unequivocally established the serious mental health implications of witnessing violence. Although there are no sufficient data from other war-torn countries or displaced populations to draw comparisons, the rate of exposure of our sample seems quite high. By comparison, Macksoud and Aber (1996), using a similar method, reported that, on average, Lebanese children ($N = 224$) in four demographically distinct areas had experienced six different types of violent incidents. The findings of this study thus speak about the moral realities and moral judgments of children who had grown up in the midst of severe political violence.

The first set of assessments ("general assessments") indicated that displaced children and adolescents think about matters of welfare and justice in moral terms. Indeed, participants in this study made universal and noncontingent moral judgments not unlike those of normative samples studied in the United States and other countries (Smetana, 2006; Wainryb, 2006). They thought it was wrong to steal and hurt others not because one may get punished or because it is against the law but because of considerations with justice and the welfare of others. Nearly all also judged it would be wrong to steal or inflict harm even if it were legal or common. No differences in these regards were found based on age or gender. These findings, that war-affected children and adolescents display noticeable moral knowledge in spite of having been exposed to violence, poverty, and dislocation, are incompatible with the grim picture of moral disorientation and truncated development painted by some (Fields, 1973, 1980). It appears that rather than merely taking their common experiences of violence and injustice as the sole criterion for determining what is right and wrong, war-affected children, not unlike children living in nonviolent communities across the world, make judgments in relation to what they perceive to be the intrinsic features of moral violations. Indeed, even common practice and the absence of laws do not—in their judgment—alter what is morally right and wrong. In this respect, these findings are consistent with an emerging body of data indicating that children develop prescriptive and generalizable moral concepts even in contexts that do not seem to facilitate, explicitly or implicitly, such a development. Whereas previous research (e.g., Helwig et al., 2003, 2007; Neff, 2001; Turiel & Wainryb, 1998; Wainryb & Turiel, 1994) has focused on contexts defined in terms of cultures that impede (or appear to impede) the development of certain moral concepts, the present study extends these findings to social contexts framed in terms of war and the accompanying conditions of violence, poverty, lawlessness, and displacement.

The lack of evidence of truncated development does not, however, imply that the moral development of children growing up in the midst of political violence is unaffected. Rather, the findings of this study point to the importance of examining both how war-affected children think about moral matters in general and how they bring their moral concepts to bear on specific contexts in which relevant conflicting considerations are made salient. Indeed, participants' psychological judgments about what story protagonists are likely to do and feel in specific conflict situations, and their moral judgments concerning the protagonists' moral transgressions in those situations, demonstrate that their moral concepts bear more heavily in some contexts than others.

Participants were first asked to predict what story protagonists are likely to do in the face of dire need or revenge. In spite of their negative judgments of theft and physical harm in the abstract (general assessments), 85% expected protagonists to steal in the survival condition, 85% expected them to steal or inflict harm in the revenge condition, and 45% expected protagonists to steal even in the baseline condition in which concerns with survival and revenge were not mentioned. These findings reflect, probably quite accurately, the social and interpersonal reality within which participants in our study function. People, they are telling us, steal and hurt each other even though such actions are morally wrong. Given that in responding to our queries, most participants suggested that they identified with the story protagonists (as in "If that guy were me, I'd try to take the bike and run"), these findings can also be taken to mean that participants think *they* might also steal or inflict harm on others.

In some respects, these findings are reminiscent of findings concerning girls and women in patriarchal societies, who judged that females have a right to make their own decisions and that it is wrong for fathers and husbands to curtail their daughters' and wives' freedom of choice but, at the same time, also predicted that most daughters and wives would acquiesce to the restrictions imposed on them by males in their families (Wainryb & Turiel, 1994). Arguably, as was the case with those girls and women, displaced children's acknowledgment that there are discrepancies between what should happen and what is likely to happen is both realistic and adaptive in the context of the social world within which they function. At the same time, their construal of the world as one in which people do not follow ethical principles might also have long-term implications for their ability or motivation to trust others, and indeed, themselves, to honor commitments and control their own aggression (see Garbarino & Kostelny, 1996). More research is needed to further examine these issues.

After participants stated what *they* thought the protagonists were likely to do in each condition (baseline, survival, and revenge), they were told that protagonists decided to engage in stealing and, subsequently, physical harm. Then they were asked to predict what the protagonists might feel after stealing or inflicting harm in each condition and to judge whether these actions were right or wrong. As might be expected based on research with normative samples, the large majority (86%) recognized that victims in all conditions would feel sad and, to a lesser extent, angry, and a majority also predicted that protagonists in the baseline (72%) and survival (68%) conditions would experience moral emotions (guilt or shame) or mixed emotions (a mixture of guilt or shame and happiness). These findings are important in light of the centrality attributed to children's understandings of affective reactions in the context of moral transgressions (Arsenio et al., 2006). Indeed, the fact that participants can appropriately predict the emotional states likely to ensue after stealing and hurting others—namely, that victims are negatively affected and transgressors feel remorse—bodes well for these children's capacity to bring their moral concepts to bear on at least some situations of moral transgression (Shaw & Wainryb, 2006).

Less consistent with normative data were the findings in the revenge condition. In this case, only a minority expected protagonists to experience guilt or shame (28%) or mixed emotions (18%); the remaining 54% expected protagonists to feel no moral emotions, with 33% expecting them to feel either happy or both happy and afraid or hateful. In some respects, this finding is a source for concern, as research in the United States has shown that "happy victimizer" attributions among older children and adolescents are predictive of behavioral problems and even psychopathology (Arsenio et al., 2006). It is nevertheless possible that this finding has different meanings, and thus different psychological consequences, for children in this sample. It is hard to tell what it might mean, for children who have suffered serious injury at the hands of a certain group of people, to expect not to feel remorse—or even to feel happy—upon inflicting (or thinking about inflicting) harm on them. Furthermore, it is important to recall that, in the present study, the happy victimizer attributions were restricted to the revenge condition; the very same children expected protagonists to feel shame or guilt in other conditions. It is thus possible that the consequences of this phenomenon may not be identical to those presumed to ensue among juvenile delinquents or aggressive children in the United States, for whom such attributions are more generalized.

As was the case for their expectations about what protagonists might feel upon hurting others, participants' moral judgments of protagonists' actions also varied by condition. As hypothesized, nearly all (99% – 100%) judged such actions to be morally wrong in the baseline condition, where no conflicting concerns were present. More important, and somewhat unexpectedly, nearly all (98% – 99%) also judged that it was morally wrong to steal or inflict harm on others in the survival condition, in which protagonists' dire economic needs were underscored. These judgments are particularly significant given that most participants also recognized the urgency of the underlying needs and expected that, in this condition, people would actually steal. Still, their judgments were unequivocally negative and moral—that is, they thought that it would be wrong for protagonists to steal or hurt others even if these acts were to fulfill some of their basic needs, not because protagonists might be caught or punished for doing so, but because of concerns with fairness and the well-being of the victims. The fact that their expectation that most people *would* steal did not translate into their judging that doing so is right, suggests that participants do not merely adhere to a relativistic view that "anything goes." It is furthermore unlikely that this finding—that they overwhelmingly judged stealing in the survival condition to be morally wrong—may be the result of social desirability; as seen below, their judgments of stealing and harming in the revenge condition were neither straightforwardly negative nor systematically grounded on moral terms.

Indeed, in the revenge condition, 29% and 24%, respectively, for stealing and harming, judged that it was okay, or partially okay, for protagonists to engage in such actions (e.g., "I think it's okay that he beat him up because that person had hurt his family first so he's just paying him back"), and an additional 39% stated that these actions were wrong but not for moral reasons, but rather because protagonists might be hurt or punished for engaging in said actions (e.g., "I don't think it was okay that he hit him and kicked him, you know, because then those people can come back and hurt his family all over again and even worse").

These findings, concerning participants' endorsement of harm in the context of revenge or their disapproval of revenge for selfish, rather than moral, reasons, are inconsistent with findings obtained among typically developing children in the United States. Although revenge is not a foreign concept among people in the United States, as attested, among others, by the broad support of capital punishment, in samples of school-aged children in the United States, only aggressive children tend to approve of revenge; nonaggressive children (aged 8 and older) consistently disapprove of revenge as a means for solving interpersonal problems and redressing wrongs, even when directly provoked (Astor, 1994; Smetana et al., 2003). Rather, these findings are reminiscent of early moral development findings concerning the potentially injurious effects of a depressed moral atmosphere, such as that of prison settings, where reasoning stabilizes around Stage 2 "tit-for-tat" morality (Kohlberg, Scharf, & Hickey, 1971; see also Gibbs, 2003). Recent research similarly suggests that, in sociopolitical contexts characterized by deprivation and neglect, children approve of the exclusion of certain people and endorse revenge (Ardila-Rey, Killen, & Brenick, in press; Brenick et al., 2007; Killen, Henning, Kelly, Crystal, & Ruck, 2007; Margie, Brenick, Killen, Crystal, & Ruck, 2006). It should be noted that, unlike what is assumed under the "moral atmosphere" hypothesis, our own findings as well as the more recent findings emerging from Killen's lab indicate that children's endorsement of retaliation and retribution is restricted to certain conditions. Even when thus limited, however, the construal of aggression against certain people as a legitimate means for redressing grievances and reinstating justice may be seen as a reason for concern as it is likely to contribute to the perpetuation of cycles of violence (Garbarino & Kostelny, 1993; Opotow, 1990; Opotow, Gerson, & Woodside, 2005; Wainryb & Pasupathi, in press).

The potential vulnerabilities identified in participants' thinking were more pronounced among adolescents. In fact, on nearly each dimension tested, adolescents fared more poorly than children. Adolescents predicted more stealing in the baseline condition (when considerations of need and revenge went unmentioned) and more violence against persons in the revenge condition, they referred more often to retribution and hate as reasons for expecting violence, they attributed more hate and happiness and less guilt and shame to perpetrators in the revenge condition; and they had a slightly higher tendency to be accepting of stealing and inflicting physical harm in the revenge condition. It is important to emphasize, however, that, as would be expected from a reliable body of moral judgment findings (Smetana, 2006; Turiel, 1998), children and adolescents did not differ in their *moral* concepts. Rather, concerns with retribution and revenge, in particular, seemed to weigh more heavily on adolescents' expectations and judgments. This finding may be thought to be due to adolescents' generalized tendency to focus on peer groups and loyalty or to their presumably more relativistic orientation that might lead them to attribute to others greater self-interest. However, developmental research conducted in a variety of communities not exposed to political violence, in the United States and other countries, has clearly demonstrated that adolescents do not endorse revenge (Smetana et al., 2003) or the exclusion of individuals or groups (Killen, Lee-Kim, McGlothlin, & Stangor, 2002), nor do they display a generalized relativistic orientation (Wainryb, 1993; Wainryb et al., 2001). This would suggest that the age-related findings in this study are more likely to be associated to the specific experiences of war-displaced adolescents than to broader features of adolescent thinking. As shown, repeated and prolonged exposure to violence does not protect children against negative developmental outcomes but, rather, increases their susceptibility to traumatization (e.g., Garbarino & Kostelny, 1996). Given the length of the political conflict in Colombia (over 50 years), it is reasonable to assume that adolescents had been exposed to these conditions longer than children.

In contrast to age, gender had a negligible effect. In the present sample, boys and girls did not differ in the level of violence to which they had been reportedly exposed (though boys reported a slightly higher level of victimization) or in their moral thinking. Although there are data suggesting that boys are typically exposed to more violence than girls (Wright et al., 1998), these data come largely from research conducted in inner-city neighborhoods in the United States. It may be that in situations of displacement following political violence, girls do not benefit from any significant additional protection. (There is, in fact, evidence indicating that in Colombia, girls comprise nearly half of the guerrilla units; Human Rights Watch, 2003). Although these gender-related findings cannot be taken to imply that all their experiences were necessarily identical, at least in regard to the dimensions assessed in this study, boys and girls displayed similar moral understandings and vulnerabilities.

Overall, this study's findings are not suggestive of moral disorientation or truncated development. Rather they speak about how war-affected children—children for whom violence, injustice, and dislocation are normalized, a fact of life—interpret and reason about distinct morally relevant contexts. In general, these findings unveil a reservoir of moral knowledge among these children: It appears that even the impoverished environments of war and displacement present youths with opportunities for reflecting on the intrinsic features of actions that harm others. At the same time, these findings also point to potential vulnerabilities in these children's moral lives. Concerns with survival—whose gravity participants clearly acknowledged—prevailed over moral considerations in some respects: Participants expected people to steal when faced with overwhelming need but also judged that doing so would be morally wrong. Concerns with revenge colored their expectations and moral judgments more sweepingly. It seems therefore possible that contexts underscoring concerns with survival might compromise children's ability to view themselves and others as moral agents, whereas contexts underscoring revenge might give rise to cycles of violence; future research might explore further these specific vulnerabilities. The long-term developmental implications of this pattern of vulnerabilities might also hinge on how children make sense, as a whole, of the political conflict within which their moral lives, and unique moral conflicts, are embedded. Societies tend to superimpose, on the reality of violence, different kinds of meanings—more or less coherent, more or less polarized—about, for example, the roots of the conflict and the nature of the "enemy." Whether such ideological frameworks serve as a protective factor vis-a-vis children's moral development (Barber et al., 2006) or foment further vulnerability (Cairns, 1996; Garbarino & Kostelny, 1993; Punamäki, 1996; Wainryb & Pasupathi, in press; Wessells, 2006) is another question that merits further investigation.

References

Ardila-Rey, A., Killen, M., & Brenick, A. (in press). Displaced and non-displaced Colombian children's reasoning about moral transgressions, retaliation, and reconciliation. *Social Development*.

Arsenio, W. F., Gold, J., & Adams, E. (2006). Children's conceptions and displays of moral emotions. In M. Killen & J. G. Smetana (Eds.), *Handbook of moral development* (pp. 581–609). Mahwah, NJ: LEA.

Astor, R. (1994). Children's moral reasoning about family and peer violence: The role of provocation and retribution. *Child Development*, *65*, 1054–1067.

Barber, B. K., Schluterman, J. M., Denny, E. S., & McCouch, R. J. (2006). Adolescents and political violence. In M. Fitzduff & C. Stout (Eds.), *The psychology of resolving global conflicts: From war to peace, Vol. 2: Group and social factors* (pp. 171–190). Westport, CT: Praeger.

Brenick, A., Killen, M., Lee-Kim, J., Fox, N., Leavitt, L., Raviv, A., et al. (2007). *Social understanding in Israeli-Jewish, Israeli-Palestinian, Palestinian, and Jordanian 5-year old children: Moral judgments and stereotypes*. Unpublished manuscript, University of Maryland, College Park.

Buka, S. L., Stichick, T., Birdthistle, I., & Earls, F. (2001). Youth exposure to violence: Prevalence, risks and consequences. *American Journal of Orthopsychiatry, 71*, 298–310.

Cairns, E. (1996). *Children and political violence*. Cambridge, MA: Blackwell.

Cairns, E., & Dawes, A. (1996). Children: Ethnic and political violence—a commentary. *Child Development, 67*, 129–139.

Davidson, P., Turiel, E., & Black, A. (1983). The effect of stimulus familiarity on the use of criteria and justifications in children's social reasoning. *British Journal of Developmental Psychology, 1*, 49–65.

Dawes, A. (1994). The effects of political violence on socio-moral reasoning and conduct. In A. Dawes & D. Donald (Eds.), *Childhood and adversity* (pp. 200–219). Cape Town, South Africa: David Philip.

Fields, R. M. (1973). *A society on the run: A psychology of Northern Ireland*. Baltimore, MD: Penguin.

Fields, R. M. (1980). *Northern Ireland: Society under siege*. New Brunswick, NJ: Transaction Books.

Garbarino, J., & Kostelny, K. (1993). Children's response to war: What do we know? In L. A. Leavitt & N. A. Fox (Eds.), *The psychological effects of war and violence on children* (pp. 23–40). Hillsdale, NJ: LEA.

Garbarino, J., & Kostelny, K. (1996). The effects of political violence on Palestinian children's behavior problems: A risk accumulation model. *Child Development, 67*, 33–45.

Gibbs, J. C. (2003). *Moral development and reality: Beyond the theories of Kohlberg and Hoffman*. Thousand Oaks, CA: Sage.

Helwig, C. C. (1995). Adolescents' and young adults' conceptions of civil liberties: Freedom of speech and religion. *Child Development, 66*, 152–166.

Helwig, C. C., Arnold, M. L., Tan, D., & Boyd, D. (2003). Chinese adolescents' reasoning about democratic and authority-based decision making in peer, family, and school conflicts. *Child Development, 74*, 783–800.

Helwig, C. C., Arnold, M. L., Tan, D., & Boyd, D. (2007). Mainland Chinese and Canadian adolescents' judgments about the fairness of democratic and other forms of governments. *Cognitive Development, 22*, 96–109.

Human Rights Watch. (2003). *"You'll learn not to cry:" Child combatants in Colombia*. New York: Author.

Human Rights Watch. (2006). *Colombia: Displaced and discarded*. Retrieved March 22, 2006, from http://hrw.org/reports/2005/colombia1005/7.htm

Killen, M., Henning, A., Kelly, M. C., Crystal, D., & Ruck, M. (2007). Evaluations of interracial peer encounters by majority and minority U.S. children and adolescents. *International Journal of Behavioral Development, 31*, 491–500.

Killen, M., Lee-Kim, J., McGlothlin, H., & Stangor, C. (2002). How children and adolescents evaluate gender and racial exclusion. *Monographs of the Society for Research in Child Development, 67*(4, Serial No. 271).

Kohlberg, L., Scharf, P., & Hickey, J. (1971). The justice structure of the prison: A theory and an intervention. *Prison Journal, 51*, 3–14.

Kupek, E. (2006). Beyond logistic regression: Structural equations modeling for binary variables and its application to investigating unobserved confounders. *BMC Medical Research Methodology, 6*, 13.

Leavitt, L. A., & Fox, N. A. (1993). Introduction. In L. A. Leavitt & N. A. Fox (Eds.), *The psychological effects of war and violence on children* (pp. xi–xix). Hillsdale, NJ: LEA.

Macksoud, M., & Aber, L. (1996). The war experiences and psychological development of children in Lebanon. *Child Development, 67*, 70–88.

Margie, N. G., Brenick, A., Killen, M., Crystal, D., & Ruck, M. (2006, May). *Peer victimization and moral judgments about interracial peer encounters*. Poster presented at the Association for Psychological Science, New York City.

Neff, K. (2001). Judgments of personal autonomy and interpersonal responsibility in the context of Indian spousal relationships: An examination of young people's reasoning in Mysore, India. *British Journal of Developmental Psychology, 19*, 233–257.

Opotow, S. (1990). Moral exclusion and injustice. *Journal of Social Issues, 46*, 1–20.

Opotow, S., Gerson, J., & Woodside, S. (2005). From moral exclusion to moral inclusion: Theory for teaching peace. *Theory into Practice, 44*, 303–318.

Punamäki, R. L. (1996). Can ideological commitment protect children's psychosocial well-being in situations of political violence? *Child Development, 67*, 55–69.

Rafaelli, M., & Larson, R. W. (1999). *Homeless and working youth around the world: Exploring developmental issues. New Directions for Child and Adolescent Development*. San Francisco: Jossey Bass.

Red Bogotá. (2006). *Hacer público lo público*. Retrieved March 10, 2006, from http://www.redbogota.com/lopublico/index.htm

Richters, J. E., & Martinez, P. (1993). Children as victims of and witnesses to violence in a Washington, D.C. neighborhood. In L. A. Leavitt & N. A. Fox (Eds.), *The psychological effects of war and violence on children* (pp. 243–280). Hillsdale, NJ: LEA.

Selner-O'Hagan, M. B., Kindlond, D. J., Buka, S. L., Raudenbush, S. W., & Earls, F. J. (1998). Assessing exposure to violence in urban youth. *Journal of Child Psychology and Psychiatry, 39*, 215–224.

Shaw, L. A., & Wainryb, C. (2006). When victims don't cry: Children's understandings of victimization, compliance, and subversion. *Child Development, 77*, 1050–1062.

Smetana, J. G. (2006). Social-cognitive domain theory: Consistencies and variations in children's moral and social judgments. In M. Killen & J. G. Smetana (Eds.), *Handbook of moral development* (pp. 119–154). Mahwah, NJ: LEA.

Smetana, J. G., Campione-Barr, N., & Yell, N. (2003). Children's moral and affective judgments regarding provocation and retaliation. *Merrill-Palmer Quarterly, 49*, 209–236.

Straker, G., Mendelson, M., Moosa, F., & Tudin, P. (1996). Violent political contexts and the emotional concerns of township youth. *Child Development*, *67*, 46–54.

Turiel, E. (1998). The development of morality. In W. Damon & N. Eisenberg (Eds.), *Handbook of psychology* (5th ed., Vol. 3): N. Eisenberg (Ed.), Social, emotional and personality development (pp. 863–932). New York: Wiley.

Turiel, E., & Wainryb, C. (1998). Concepts of freedoms and rights in a traditional hierarchically-organized society. *British Journal of Developmental Psychology, 16*, 375–395.

United Nations. (2006). *Special representative of the Secretary-General for Children and Armed Conflict: The impact of armed conflict on children*. Retrieved March 10, 2006, from http://www.un.org/special-rep/children-armed-conflict/index.html

Wainryb, C. (1993). The application of moral judgments to other cultures: Relativism and universality. *Child Development*, *64*, 924–933.

Wainryb, C. (2004). Is and ought: Moral judgments about the world as understood. In B. Sokol & J. Baird (Eds.), *New directions for child and adolescent development: Vol. 103. Mind, morals, and action: The interface between children's theories of mind and socio-moral development* (pp. 3–17). San Francisco: Jossey-Bass.

Wainryb, C. (2006). Moral development in culture: Diversity, tolerance, and justice. In M. Killen & J. G. Smetana (Eds.), *Handbook of moral development* (pp. 211–240). Mahwah, NJ: LEA.

Wainryb, C., & Brehl, B. (2006). "I thought she knew that would hurt my feelings": Developing psychological knowledge and moral thinking. In R. Kail (Ed.), *Advances in child development and behavior* (Vol. 34, pp. 131–171). Amsterdam: Academic Press.

Wainryb, C., Brehl, B., & Matwin, S. (2005). Being hurt and hurting others: Children's narrative accounts and moral judgments of their own interpersonal conflicts. *Monographs of the Society for Research in Child Development, 70*(Serial No. 281).

Wainryb, C., & Pasupathi, M. (in press). Developing moral agency in the midst of violence: Children, political conflict, and values. In A. Karawan, W. McCormack, & S. E. Reynolds (Eds.), *Values and violence: Intangible aspects of terrorism*. New York: Springer.

Wainryb, C., Shaw, L. A., Laupa, M., & Smith, K. R. (2001). Children's, adolescents', and young adults' thinking about different types of disagreements. *Developmental Psychology, 37*, 373–386.

Wainryb, C., & Turiel, E. (1994). Dominance, subordination, and concepts of personal entitlements in cultural contexts. *Child Development*, *65*, 1701–1722.

Wessells, M. (2006). Child soldiering: Entry, reintegration, and breaking cycles of violence. In M. Fitzduff & C. Stout (Eds.), *The psychology of resolving global conflicts: From war to peace*, Vol. 3: Interventions (pp. 243–266). Westport, CT: Praeger.

World Bank (2000). *Violence in Colombia: Building sustainable peace and social capital*. Washington, DC: Author.

Wright, M. O., Master, A. S., & Hubbard, J. J. (1998). Long-term effects of massive trauma: Developmental and Psychobiological perspectives. In D. Cicchetti & S. L. Toth (Eds.), *Developmental perspectives on trauma: Theory, research, and intervention* (pp. 181–225). Rochester, NY: University of Rochester Press.

Child Development, July/August 2008, Volume 79, Number 4, Pages 899–913

Pathways to Literacy: A Study of Invented Spelling and Its Role in Learning to Read

Gene Ouellette
Mount Allison University

Monique Sénéchal
Carleton University

This intervention study tested whether invented spelling plays a causal role in learning to read. Three groups of kindergarten children (mean age = 5 years 7 months) participated in a 4-week intervention. Children in the invented-spelling group spelled words as best they could and received developmentally appropriate feedback. Children in the 2 comparison groups were trained in phonological awareness or drew pictures. Invented-spelling training benefited phonological and orthographic awareness and reading of words used in the intervention. Importantly, the invented-spelling group learned to read more words in a learn-to-read task than the other groups. The finding are in accord with the view that invented spelling coupled with feedback encourages an analytical approach and facilitates the integration of phonological and orthographic knowledge, hence facilitating the acquisition of reading.

Young children often attempt to represent words in print before they can reliably read words. When children first begin experimenting with the written code, their spelling attempts may bear little resemblance to conventional spellings. Over time, these invented spellings increase in phonological accuracy and orthographic complexity as children become more adept at capturing words in print. Given that invented spelling most often emerges before children are established readers (Gentry & Gillet, 1993), it is possible that these early spelling attempts promote the development and integration of the skills that facilitate the acquisition of reading. At present, however, there is limited experimental evidence that invented spelling is indeed a valuable precursor to early word reading. The goal of the present research was to directly assess the role of invented spelling in early reading acquisition. A better understanding of the causal relation between early spelling and reading can contribute to models of early literacy acquisition as well as empirically validated approaches to early literacy education.

Children's early attempts to represent words in print prior to any direct instruction in spelling have been of interest to researchers and clinicians ever since the topic was brought to the forefront by Read (1971, 1975), Chomsky (1971), and Clay (1972), and subsequently discussed in detail by Treiman (1993). Read described how as children experimented with putting words into print, they became more adept at connecting the sounds heard to the letters produced. Hence, over time, invented spellings increase in sophistication and gradually come to resemble conventional spellings. Subsequent descriptions have revealed a progression from initial nonalphabetic markings to capturing the initial sounds, then to the initial and final sounds, and then to the appearance of medial vowels in children's invented spellings (Ferreiro, 1991; Gentry & Gillet, 1993; Henderson, 1981; Treiman, 1993; also see in Hebrew and Dutch: Levin, Both-de Vries, Aram, & Bus, 2005; Levin & Korat, 1993). In accordance with Read's original descriptions, subsequent stage theories of invented spelling have also proposed that invented spelling progresses from an initial reliance on phonology in translating what is heard into print to a later incorporation of word-specific orthography (e.g., Gentry & Gillet, 1993). Key to the concept of invented spelling is that it results from a child's own experimentation with representing words in print and it occurs naturally (Burns & Richgels, 1989; Richgels, 1995). It is not a process of memorization and recall of conventional spellings; rather, it is a developmental progression in which spelling attempts increase in phonological and orthographic accuracy over time.

This research was supported by a grant from the Social Sciences and Humanities Research Council of Canada to the second author. We thank the children, parents, teachers, and school personnel who made this study possible. We also thank Tina Leclaire and Kristen Petruska for their invaluable help with data collection.

Correspondence concerning this article should be addressed to Gene Ouellette, Department of Psychology, Mount Allison University, 49A York Street, Sackville, New Brunswick, Canada E4L 1C7. Electronic mail may be sent to gouellette@mta.ca.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7904-0007

The issue addressed in the present research is whether invented spelling allows children to develop an analytic stance to language that goes beyond phonological awareness stimulation and promotes the storage of orthographic information within lexical representations, beyond learning letter names and sounds. There has been abundant research showing that children benefit more from phonological awareness training when the intervention also includes letter knowledge training as compared to when it does not. In their meta-analytic review, Ehri et al. (2001) found 18 different phonological awareness training studies that included a letter knowledge training component and were conducted with normally developing or at-risk children in kindergarten or Grade 1. Ehri et al. reported that the overall effect sizes for these intervention studies were statistically significantly greater than those for studies that did not include a letter knowledge training component (i.e., the reported effects for the training with and without letters were 1.11 vs. 0.83, 0.67 vs. 0.38, and 0.61 vs. 0.34, for outcome measures of phonological awareness, reading, and spelling, respectively). Ehri et al. argued that letter training in conjunction with phonological awareness training allows young children to anchor the abstract notion of phonemes onto the concrete symbols that are letters. We contend that invented spelling also has this facilitative effect and more. Invented spelling may indeed help children analyze oral language into phonemes, but, very importantly, it provides an exploratory learning experience that may also promote the integration of phonological and alphabetic information into initial lexical representations. As these lexical representations become refined, they may facilitate the acquisition of reading. The importance of integrating phonology and orthography in reading acquisition is evident in the lexical quality hypothesis of Perfetti and Hart (2002), which attributes success in early word reading to the *quality* of internal lexical representations, as defined by integrated knowledge of a word's phonology, orthography, and meaning. Invented spelling, therefore, may allow children to explore and analyze the written code and begin to make these proposed important associations between phonological and orthographic representations.

In a similar vein, Frith (1985) proposed that early spelling attempts provide children with insight into the role of the alphabet in representing sounds in words, and this knowledge is later transferred to word reading. In fact, when invented spelling is included in empirical investigations, correlations suggest that invented spelling predicts early reading (Mann, 1993; McBride-Chang, 1998; Morris & Perney, 1984; Shatil, Share, & Levin, 2000). Further support for a pathway of influence from invented spelling to word reading comes from a word learning study conducted by Richgels (1995). In this experiment, kindergarten children with strong alphabetic knowledge but who were still nonreaders were classified as good or poor invented spellers. These children (n = 16) were then taught to read phonetically simplified words through a paired-associate learning task. The good invented-spelling group significantly outperformed the poor invented-spelling group in terms of the number of words learned over a 2-day period. Richgels concluded that there is a strong relation between invented spelling and learning to read at the start of schooling.

Training studies offer stringent tests of a causal relation between invented spelling and early word reading. That is, if training in one area can be shown to transfer to or facilitate growth in another, hypothesized pathways of influence can be confirmed. Following the proposed directional influence of invented spelling on word reading, training invented spelling should facilitate the acquisition of word reading (Frith, 1985; Treiman, 1998). Two studies are frequently cited as evidence that training invented spelling may improve reading: Ehri and Wilce (1987) and Uhry and Shepherd (1993). Ehri and Wilce found that kindergarten children trained to spell phonetically simplified words (e.g., *seat* spelled as *set*) learned to read similarly simplified words in fewer trials than did a control group of children. In this study, however, instruction was not tailored to the individual child's developmental level, nor did training involve spontaneous spelling attempts that are fundamental to the definition of invented spelling. In addition, the experimental group may have been more adept at learning to read simplified words because they were more familiar with similar stimuli. That is, participants in the experimental group were exposed to simplified spellings that were taught as words during the training component, whereas the control group was not.

The training program reported by Uhry and Shepherd (1993) involved teaching Grade 1 students to phonologically analyze words: A segmentation task was taught and letters incorporated as an introduction to *conventional* spellings. Impressively, trained children demonstrated transfer to word decoding and outperformed a control group of children on posttest measures of word reading. Similar results were reported by O'Connor and Jenkins (1995) with a small sample of children who had developmental delays. Although these studies highlight the importance of analytic phonological awareness and conventional spelling instruction in Grade 1, they did not directly deal with the potential pathway of influence between *invented* spelling at the start of schooling and subsequent word reading. In fact, an extensive review of the literature revealed only two training studies in which naturally occurring invented spelling was encouraged or involved in a teaching paradigm and word reading was evaluated as a posttest measure. These two studies are outlined next.

In a naturalistic educational study, Clarke (1988) compared two Grade 1 classroom programs—one that included invented spelling and one that included conventional spelling instruction. Over the school year, the classes in which invented spelling was encouraged showed the greatest improvements in both decoding and irregular word recognition. This important study establishes the relevance of invented spelling in the classroom, yet the natural classroom context of the research comes at the expense of experimenter control and internal validity. In addition, phonological awareness was not measured, and thus, it is not known if group differences were attributable to any preexisting differences in this already established influential skill.

Rieben, Ntamakiliro, Gonthier, and Fayol (2005) recently reported a carefully controlled multiple-group intervention study with 5-year-old French-speaking children. This study contrasted the effects of training invented spelling, invented spelling with corrective feedback, and copied spelling. The invented-spelling group with feedback scored significantly higher than the control group and all other experimental groups for reading practiced words. However, no between-group differences were evident for reading of words not used in the training program, and more surprisingly, no group differences were reported for measures of phonological awareness and invented-spelling complexity. The limited impact of invented spelling may be due to the fact that the 18 training sessions given to each group were spread out over 6 months, which may not have been intensive enough to promote learning. Participants' letter knowledge was also very low, which may have constrained learning across all conditions and precluded those in the invented-spelling groups from benefiting from the intervention, especially given the complexity of the reading test stimuli used.

The research reviewed so far provides some evidence that early spelling attempts influence subsequent word reading, but strong experimental evidence is still lacking (Caravolas, Hulme, & Snowling, 2001; Fletcher-Flinn, Shankweiler, & Frost, 2004; Rieben et al., 2005). The present study evaluated the benefits of training kindergarten students to be better invented spellers through a pretest, posttest, comparison group design. In addition to a group trained in invented spelling and a control group, there was also a comparison group trained in phonological awareness. Given the meta-analytic results obtained by Ehri et al. (2001), all children, including those in the phonological awareness group, also received letter knowledge training. The inclusion of a phonological awareness (plus letters) group allowed for a comparison between children trained on invented spelling and those trained on phonological awareness, a comparison that is relevant, given past assertions that invented spelling represents self-directed phonological awareness training (e.g., Mann, Tobin, & Wilson, 1987). This design allowed for a clear evaluation of group differences on key skills that were posttested, following the different interventions. In addition, a learn-to-read task was given to all participants following the intervention to assess any facilitative effects of the training on *learning* to read words. All participants were children who, at the onset of the intervention study, did not demonstrate consistent word reading ability. This allowed us to test the transfer of skills from spelling to reading without the influence of well-established, preexisting reading skills.

It was hypothesized that the training procedures implemented would prove to be efficient in that the invented-spelling group would show improved spelling sophistication relative to the other groups, whereas the phonological awareness group would show improved phonological awareness on posttest relative to the control group. Given that invented-spelling practice may involve phonological awareness, it was hypothesized that the invented-spelling group would also show a posttest advantage over the control group for phonological awareness. It was also hypothesized that the invented-spelling group would show a posttest advantage over both other groups on a task of implicit orthographic awareness. Central to the study, the invented-spelling group was hypothesized to outperform the other groups on a learn-to-read task. Although we also included a general word reading posttest measure, we did not think that a 4-week intervention would be long enough to show the transfer from invented spelling to word reading without some form of reading instruction.

Method

Participants

Participants were selected to meet a number of criteria from a sample of 145 kindergarten children who were participating in another research project (Ouellette & Sénéchal, 2008). All children used English as their primary language and were recruited from four English schools in a large Canadian city. To control for the influence of prior reading skills, children who could read more than one word from a list of five simple consonant-vowel-consonant decodable words and/or more than one word from a list of five high-frequency sight words were not considered for the present study (n = 30).

From those identified as having no or minimal word reading ability, children from each participating class were randomly assigned to one of three intervention groups if they could be matched with two other students from the same class on invented spelling, letter–sound knowledge, and phonological awareness. In all, this matching procedure resulted in 69 children (39 female and 30 male, mean age = 5 years 7 months) assigned to one of three intervention groups, creating three equivalent groups of 23 students. The mean level of education of the parent completing the permission form was some postsecondary studies (college or undergraduate courses), and parental education ranged from *not having completed high school* to *having completed postgraduate studies*. The sample was chiefly Caucasian, yet multicultural in the sense that 43.5% reported having a second language spoken in the home, which included French (27.5%), Asian (4.3%), and other European languages (11.7%). Because the matching was done within each class, classroom/teacher effects were counterbalanced across groups. In all, children were drawn from nine classes from four schools. In their kindergarten year, children in this Canadian city only attend school for half of the day, had not received direct phonics instruction beyond exposure to the alphabet, and their teachers did not follow a standard kindergarten literacy curriculum.

Measures

Invented spelling. Children were asked to spell as best they could 10 words. The words were all selected to meet a certain number of criteria. The words could be spelled by relying exclusively on a set of 13 letters; they were high-frequency words, contained relatively high-frequency biphone probabilities, and represented a variety of articulatory and syllabic features. Three of these words were taken from Tangel and Blachman (1992, 1995), whereas the remaining words were chosen to fulfill several criteria and are presented in the Appendix.

The scoring system described by Tangel and Blachman (1992, 1995) was used to evaluate the invented-spelling attempts because it is sensitive to the developmental sequence observed in descriptive studies. Each spelling attempt was scored on a 7-point scale (0–6). The scale takes into account both the number of phonemes represented and the level of orthographic representation and has excellent reported reliability. Typically, a score of 0 is given for a random string of letters, a score of 1 for a phonetically related letter used to mark a salient part of the word, a score of 2 for the correct initial grapheme, a score of 3 for representing more than one phoneme with phonetically related or conventional letters (in the proper sequence), a score of 4 for all phonemes represented with phonetically related or conventional letters, a score of 5 for all consonant phonemes represented with conventional letters and vowel presence marked, and a score of 6 for proper conventional spelling.

Interrater reliability was initially assessed by having two raters score all spelling attempts (Tangel & Blachman, 1992, 1995; Treiman & Bourassa, 2000). The percent agreement between the two raters on all 1,150 items (115 eligible children × 10 words) was 85.83%, and the Pearson correlation between the total score given to each child by the two raters was strong at r = .98. Each disagreement was reviewed by the raters until a consensus score reached. This is an important step to ensure accuracy, as 119 of 163 discrepancies were found to be scorer error, whereas the remaining 44 discrepancies were related to interpretation of the child's intended target letter. Using the final agreed-upon scores, the interitem reliability was also calculated and found to be excellent (α = .91).

Word reading. At pretest, children were asked to read 10 words to assess word reading. The word list included 5 high-frequency, irregularly spelled sight words taken from lists provided by both Fry, Kress, and Fountoukidis (2000) and Fountas and Pinnell (1996), and 5 high-frequency, decodable words. The 10 words were between two and four letters in length and are presented in the Appendix. The decodable words were also chosen based on their concreteness so that they could be easily drawn, as well as to represent various syllable structures. Words were presented one at a time in a fixed random order on separate index cards in 48-point font. The irregular words were shown first and the children asked to say what they thought the word was. For the decodable items, the children were encouraged to look at the letters, to sound them out, and to give their best guess. At posttest, 5 additional words were added. Note that the 5 decodable words from the reading pretest were stimuli used in the interventions; thus, as a posttest, they offer a measure of regular (decodable) word reading for practiced words. Five additional decodable words were added at posttest to provide a measure of regular word reading for novel words. These additional words were of the same syllable structure, composed from the same limited letter set, and contained the same vowel types as the practiced words (Appendix). The interitem reliability at posttest was very good (α = .78).

Letter–sound knowledge. Children were asked to provide the sounds made by 24 letters of the alphabet (all letters of the alphabet, except for X and Q) in addition to three digraphs (CH, SH, and TH). Each grapheme (or digraph) was printed on an individual index card in both upper and lower case 72-point Arial font. Cards were presented in a fixed random order, and the maximum score was 27. Interitem reliability was excellent (α = .87 at posttest).

Phonological awareness. Three subtests of the Comprehensive Test of Phonological Processing (CTOPP; Wagner, Torgesen, & Rashotte, 1999) were administered to assess children's awareness of phonemes in the word-initial and word-final positions, as well as their ability to delete a syllable or a phoneme from a spoken word, and to blend individually presented phonemes into words. In the Sound Matching subtest, children were shown four pictures on each trial and asked to indicate which picture either started the same or ended the same as the first picture. Each picture was also named by the experimenter upon presentation. There are 20 items in all (10 first sound matching and 10 final sound matching), but testing is stopped when the child misses 4 of 7 items. In the Elision subtest, the child repeats a word after the experimenter and is then asked to say the word again but leaving certain sounds out. For example, the child is asked to say cup without saying /k/. There are 20 items in all, with testing stopped following three consecutive errors. The Blending Words subtest requires the participant to listen to a recorded CD of a female voice, saying words phoneme by phoneme. The child is then asked to say what the intended word was. This subtest was presented via a personal CD player and headphones. There are 20 items in all, with testing stopped following three consecutive errors. Following the recommended scoring procedures of the CTOPP, scores on these subtests were converted to standard scores with a mean of 10 and then combined to create a phonological awareness composite. The reported reliability of this composite is very good (α = .79).

Orthographic awareness: Permissible sequences. Following the conceptualization of orthographic awareness encompassing visual recognition of permissible sequences within words (Levy, Gong, Hessels, Evans, & Jared, 2006), children's implicit awareness was measured with two forced-choice tasks. The tasks tapped implicit knowledge of rules because explicit knowledge of orthographic constraints is not expected at this age, nor would the children be expected to be able to verbalize these rules. The first task, referred to as *Vowels*, was adapted from Levy et al. (2006) and contrasted nonwords containing a medial vowel with nonwords lacking a vowel (10 nonword pairs). The procedure of Levy et al. was followed whereby each card containing a stimuli pair was shown and the child then asked which item looked best or which item their teacher would like to read the most. All stimuli were presented in a fixed random order on individual index cards in 72-point Arial capitalized font. The final task, named *Double Consonants*, was adapted from Cassar and Treiman (1997) and contrasted nonwords where one item contained an allowable consonant doubling at the final position and the other contained an illegal consonant doubling in the onset position. All stimuli are presented in the Appendix in the order of presentation. Performance here was close to chance for many of the children, thus preventing meaningful estimation of reliability.

Learn-to-read task. Posttesting concluded with a learn-to-read task modeled after that developed by Ehri and Wilce (1987). In this learn-to-read task, each participant was taught to read 10 decodable words in a paired-associate fashion. The words to be learned were all composed from the same letter set taught to all intervention groups and are listed in the Appendix. Five words started with *B* and 5 with *P*, thus requiring the children to pay attention to more than just the initial consonant. These words represent a range of common syllable types used in early reading instruction and include words of varying difficulty, with the inclusion of a consonant blend and bi-syllabic word. Children, therefore, had to attend to all letters in the words, and this small set of words allowed for their learning proficiency to be evaluated over a range of syllable types and complexity.

The procedure was adapted from that outlined by Ehri and Wilce (1987). The task began with a study trial in which children were told that they were going to learn to read some words and that the words contained the letters that they had practiced in their groups. Each word was shown to the child and read by the instructor in a sounding out and blending fashion, as he or she ran his or her finger under the word: The word was said phoneme by phoneme with the sounds stretched and still connected to encourage the children to use decoding skills, yet not disrupting coarticulatory effects. The word was then repeated at a typical speaking rate. A sentence was also read that provided a meaningful context for the word. The child was then asked to say the word in a similar fashion while running his or her finger under it. This was repeated for all the 10 words. A recall trial was then completed in which the 10 words were shown individually and the child asked to read each one by sounding out the letters; if incorrect, the answer was given and the child asked to repeat the word out loud stretching the sounds while again running his or her finger under the word. Trials (i.e., attempts at reading all 10 words) were repeated with the word order varied randomly, until all the 10 words were read correctly on two consecutive trials (or to a maximum of five trials). Reliability was excellent ($\alpha = .94$).

Intervention

The study included one invented-spelling training group and two control groups. The first control group received an alternative phonological awareness training, whereas the second control group was asked to draw pictures for the target words. All children in each of the three groups were taught 13 letter – sound associations necessary to form the training target words as well as the pre- and posttest words. Letter – sound training was included for three reasons: (a) to avoid any constraint of low letter – sound knowledge on posttest measures, (b) to optimize learning in the phonological awareness comparison group (Ehri et al., 2001), and (c) to control for a possible confounding influence of increased exposure to letters in the invented-spelling group. Letter – sound teaching was carried out through a group activity, in which the children were taught a rhythmic chant for each letter, with a sequence of clapping and knee slapping, which gave both the name and the sound of the letter with simultaneous presentation of the printed letter. This activity occurred for approximately 3 min at the start of each group session, with four letter sounds practiced each session and recycled in the same order throughout the duration of the training program.

The same target words were used in each group, in the same order, and with the same frequency of exposure. As indicated previously, all target words used during training were composed from the limited set of 13 letters taught to all groups, as were all regular words in the spelling and reading assessments (pre- and posttests). These letters included a variety of articulatory manners: voiced and voiceless stops (B, P, D, T), liquids (L, R), a nasal (N), fricative (S), glide (Y), and four common vowels (A, E, I, and O). Both short and long vowel sounds were taught.

To control for a possible confound of familiarity with syllable/word forms used in training, all groups were exposed to the same words and at the same frequency (see subsequently). There were 20 regularly spelled words used in this training study, and all were made from the 13 letters also taught to each group. The words were chosen to represent different syllable structures and vowel sounds. All words are presented in the Appendix. Note that 5 of these 20 words were also used in reading and spelling posttests, thus allowing for an evaluation of word reading and invented spelling for practiced words, respectively.

Invented-spelling group. This group was trained to increase the sophistication of their naturally occurring invented spellings. The training words were presented, one at a time, in both picture form and orally by the instructor. Each word was spoken out loud by the instructor at a normal speech rate. It was then repeated in a stretched manner with exaggerated articulation but with no pausing between the phonemes. The instructor said the word a third time (at a normal speech rate) and the children were asked to repeat the word out loud in unison. The word was said a fourth time and the children instructed to each print the word in a provided notebook. Children were instructed to print the word how they thought it would look. They were repeatedly encouraged to do their best and told that their spellings did not have to be the same as an adult might write or even be the same as the spellings produced by the other children in their group.

After each word was printed, the instructor went around the table offering individually tailored feedback in which each invented spelling was contrasted with an instructor-generated invented spelling representative of a minimal increase in sophistication. This feedback was provided in the context of praising the child's invented spelling and then showing another way to write the word; the corrected form typically contained one additional phoneme represented as per the developmental progression outlined by Gentry and Gillet (1993). For example, a child who spelled *eel* as *ekxn* would be shown an instructor-generated spelling with one additional level of sophistication— in this example the correct final phoneme. The instructor would thus copy the child's spelling, replacing the last letter with a conventional representation of the final phoneme *(ekxl)*. A similar procedure was recently used successfully with Portuguese-speaking kindergarten children by Martins and Silva (2006). Following the fourth session, feedback also included drawing the children's attention to any extra letters within their invented spellings. This was done to coincide with the phonological awareness group's switch to phoneme counting, as well as to offer additional help to the weaker spellers by focusing their attention on the correspondence between sounds within a word and the letters in their spellings. In the example mentioned previously, the corrected feedback given to the child's *ekxn* was to add the final *l* and remove the unnecessary letters, yielding *el*. Note that the procedure of providing feedback in the form of a model with one additional element of complexity resulted in providing the conventional spelling only when the child's production was one element away from being (conventionally) correct. For example, *eel* was only provided as the corrected spelling if the child produced *el* on his or her own, here, the one additional level of sophistication is to go from the accurate phonetic spelling to the conventional form. Following the individualized feedback, children turned to a clean page in their notebooks, and the procedure was repeated using the same word. After the word had been spelled (and feedback given) twice, the procedure was repeated with the next word from the list. Within each session, 5 words were spelled (twice each); these words were repeated for two consecutive sessions, yielding the total of 20 words over eight sessions; the ninth and final session repeated 5 of these words.

Phonological awareness group. This group was taught to analyze words into smaller segments. Following the developmental progression of phonological awareness tasks proposed by Ehri et al. (2001) and previous research on analytic phonological awareness training with children of this age (Ball & Blachman, 1991; Tangel & Blachman, 1992), two segmentation-related tasks were taught. Also, following previous research showing greater effects when letter–sound instruction is paired with phonological awareness training (Ehri et al., 2001), each session started with letter–sound training for the targeted sounds. Children in this group were first taught to match pictures based on shared initial and final sounds, using the first 10 of the 20 training words (in picture form), in the same order and at the same frequency as in the invented-spelling group (i.e., 5 words per session, each repeated twice and used for two consecutive sessions). Each child was given a sheet with the appropriate picture along with three others. All pictures were named by the instructor in the same fashion as for the invented-spelling group: one time at a normal speech rate, one time with stretched speech, and two more times at a normal rate, with the children repeating the word out loud in unison once. Thus, the frequency and manner of exposure to the training stimuli were carefully controlled across groups. For each target word, the children were asked to circle the pictures that started or ended the same. The instructor then went around the table offering individually tailored corrective feedback, modeling the correct answer as necessary. Each trial was then repeated.

For the final 10 words (i.e., final five sessions), the children were taught a phonemic segmenting task based on Elkonin's (1973) original *say-it and move-it* activity. In this task, children learned to represent each sound within a word by stamping a marker, once for each phoneme in a word, into squares below a picture of that word. Again, the instructor said each word four times as before, and the procedure modeled as necessary. For each word, the children were asked to repeat the word out loud in unison and to make the appropriate number of stamps as they did so. Individually tailored feedback was given and the trial repeated.

Control group. This group of children received similar instructions as the invented-spelling group, except that instead of writing each word, they were asked to draw a picture. They too heard each word four times and said each word once in unison. Following completion of each drawing, the instructor provided positive feedback to each child and then the procedure was repeated using the same word (as in the other groups). Recall that children in this control group also received letter–sound training at the beginning of each session.

Procedure

Children received the assigned intervention in small groups. Group size depended on the logistics of the number of children participating from each school and ranged from 3 to 6 children. Interventions were delivered in nine sessions scheduled over a 4-week period midway through the senior kindergarten school year. Each session lasted approximately 25 min, with equivalent instructional time given to all groups and conditions.

Interventions were delivered by one of three highly experienced instructors, including a speech language pathologist with past experience delivering intervention programs to groups of children. The other instructors were a licensed classroom teacher and an experienced research assistant. To avoid potential instructor effects, each instructor taught two invented-spelling groups, two phonological awareness groups, and two control groups (Troia, 1999).

Posttesting was completed over two individual assessment sessions, completed within 7 days of the last intervention session. Children were posttested in pseudorandom order, such that the same number of children from each intervention condition was tested each day. Posttesting included the areas hypothesized to be influenced by the types of training proposed here: letter–sound knowledge, invented spelling, phonological awareness, orthographic awareness, word reading, and word learning. All children received the same order of tasks, arranged to make for two sessions of equal length with varying task demands and response requirements.

Results

Pretest Measures

Triads of children from the same class were matched on measures of invented spelling, phonological awareness, and letter–sound knowledge. A multivariate analysis of variance (ANOVA) confirmed that there were no statistically significant group differences on the set of these measures, Wilks's lambda = .94, $F(6, 128) = 0.73$, $p = .63$. Although not a matching variable, children were also pretested on orthographic awareness, and the groups did not differ on this measure either, $F(2, 66) = 0.92$, $p = .40$. As noted, the 69 participating children were all considered to have very little or no word reading skills as indexed by their performance on a word reading screening measure (i.e., reading 10 high-frequency words). Of these children, 39 (57%) were able to read one decodable word only, 16 (23%) were able to read one irregular word only, and the remainder (20%) could not read any of the words. The intervention groups did not differ in their pretested word reading, $F(2, 66) = 0.03$, $p = .97$. The descriptive statistics for all pretest measures are presented in Table 1.

Posttest Measures

All posttest measures were entered into a multivariate ANOVA. This between-group test was statistically significant, Wilks's lambda = .65, $F(12, 122) = 2.41$, $p < .01$. Performance between groups on individual measures was then analyzed in a series of univariate ANOVA, followed by two planned orthogonal contrasts (Helmert) to address the previously stated hypotheses. With the exception of the phonological awareness analyses, the first contrast compared the invented-spelling group against the two control groups and the second compared the two control groups to each other. These analyses are presented next, followed by the results of the learn-to-read task. All posttest means and univariate test statistics are presented in Table 2.

Letter–sound knowledge. ANOVAs conducted on children's knowledge of the 13 letter sounds taught revealed no statistically significant group difference, $F(2, 66) = 0.42$, $p = .66$. There was also no difference for knowledge of the entire set of 27 graphemes/digraphs tested, $F(2, 66) = 0.35$, $p = .70$.

Invented spelling. An ANOVA on the complete invented-spelling posttest revealed a statistically significant group difference, $F(2, 66) = 3.79$, $p < .05$, with a moderate effect size ($\eta^2 = .103$). Considering the pattern of results in Table 2, the invented-spelling group appears superior to the other two groups; in support of the present hypothesis, a planned orthogonal contrast confirmed the significance of this difference, $t(66) = 2.70$, $p < .01$, whereas the other planned contrast revealed no statistically significant difference between the phonological awareness and the drawing groups, $t(66) = 0.54$, $p = .59$. Note that the same pattern of significant results was found when the analyses were repeated separately for the five words that were practiced in the intervention as well as for the five words not practiced. The higher scores observed for the invented-spelling group thus reflect the increased sophistication in the self-generated

Table 1
Means (Standard Deviations) for Each Pretest Measure

	Training group		
Measure (maximum)	Invented spelling	Phoneme awareness	Drawing
Invented spelling (60)	24.47 (11.36)	24.48 (12.65)	23.83 (10.78)
Letter sounds (27)	16.74 (6.82)	17.00 (4.41)	16.48 (5.48)
Trained letter sounds (13)	8.87 (3.98)	9.22 (2.56)	9.35 (2.95)
Phonological awareness[a]	29.35 (3.67)	30.17 (4.00)	29.39 (3.65)
Awareness of double consonants (10)	4.87 (1.58)	5.00 (2.02)	4.39 (1.92)
Awareness of vowels (10)	5.01 (1.07)	4.61 (1.44)	5.17 (1.44)
Reading: regular words (5)	0.57 (0.51)	0.61 (0.50)	0.53 (0.52)
Reading: irregular words (5)	0.22 (0.42)	0.17 (0.39)	0.30 (0.47)

[a] Comprehensive Test of Phonological Processing standard score with mean of 30.

Table 2

Means (Standard Deviations) for Each Posttest Measure and Univariate Tests for Training Groups

Measure (maximum)	Training group: Invented spelling	Training group: Phoneme awareness	Training group: Drawing	*F*
Letter sounds (27)	19.96 (4.98)	20.78 (3.03)	19.74 (5.04)	< 1
Trained letter sounds (13)	10.96 (2.12)	11.39 (1.34)	11.22 (1.76)	< 1
Invented spelling (60)	40.04 (12.01)	32.48 (13.21)	30.52 (11.89)	3.79*
Phonological awareness[a]	33.70 (5.38)	32.52 (3.93)	30.26 (4.01)	4.07*
Awareness of double consonants (10)	6.09 (2.02)	5.78 (1.57)	6.26 (2.49)	< 1
Awareness of vowels (10)	6.17 (1.23)	5.43 (1.44)	5.04 (1.55)	3.79*
Reading: practiced words (5)	2.65 (1.30)	1.65 (1.03)	1.65 (1.19)	5.52**
Reading: new words (10)	1.22 (1.35)	1.04 (1.46)	1.35 (1.75)	< 1[b]

[a]Comprehensive Test of Phonological Processing standard score with mean of 30. [b]The same pattern of nonsignificant results was found repeating this analysis of variance with only regular (decodable) novel words included or with only irregular novel words as the dependent variable.
$^*p < .05$. $^{**}p < .01$.

spellings of these children, across all words. Qualitative analyses revealed an increase in their ability to match a word's phonology with appropriate (phonetically or conventionally) letter sequences. Children in the invented-spelling group also became more adept at reducing the number of unnecessary letters in their invented spellings, as they appeared to become more aware of the correspondence between the number of phonemes heard and the number of graphemes needed. This increased sophistication is reflected in the change evident in the mean spelling score (out of 6) from 2.45 pretest to 4.00 posttest for the invented-spelling group. On average, this reflects a change from representation of the first and/or the last sound of a word to representing all sounds within an item. There was also variability evident both within groups and across words. Importantly, many children demonstrated different levels of invented spelling (i.e., as seen in variability across words), reflecting a gradual process of refinement rather than discrete stages per se, making it difficult to classify children by spelling level.

Phonological awareness. The mean performance of each group on the phonological awareness posttest composite was entered into a separate ANOVA, which revealed a statistically significant group difference, $F(2, 66) = 4.07$, $p < .05$, with a moderate effect size ($\eta^2 = .110$). To evaluate the effectiveness of the phonological awareness training, the first planned contrast here compared the phonological awareness group to the drawing group. As suggested by Table 2, the phonological awareness group did perform significantly better than the drawing group on this posttest measure, $t(66) = 2.15$, $p < .05$. The second planned contrast addressed the question of whether there would be a difference between the phonological awareness and the invented-spelling groups: No statistically significant difference was found, $t(66) = 0.85$, $p = .40$.

Orthographic awareness: Permissible sequences. An ANOVA revealed no statistically significant group difference for awareness of consonant doubling, $F(2, 66) = 0.32$, $p = .73$, but there was a significant difference for awareness of vowel presence, $F(2, 66) = 3.79$, $p < .05$, with a moderate effect size ($\eta^2 = .103$). In accordance with the present hypotheses, planned contrasts confirmed that the invented-spelling group performed significantly better than the other two groups on the vowel presence task, $t(66) = 2.59$, $p < .01$, whereas there was no statistically significant difference between the phonological awareness and the control groups, $t(66) = 0.94$, $p = .35$. The discrepancy in the pattern of results between the two orthographic awareness tasks can be reconciled with the nature of the training stimuli: Only one training word (*ladder*) contained a consonant doubling, and therefore, the children were not exposed to doublings intensively enough to promote learning of a generalized orthographic rule. Conversely, those in the invented-spelling group appear to have developed an increased awareness of the importance of vowels in spelling.

Word reading. Five regular (decodable) words in the word reading measure were also included in the training. Using these words as a dependent measure in an ANOVA revealed a statistically significant between-group difference, $F(2, 66) = 5.52$, $p < .01$, with a moderate effect size ($\eta^2 = .143$). The invented-spelling group appears superior to both the comparison groups, and planned orthogonal contrasts confirmed the significance of this difference. The invented-spelling group was found to have higher scores relative to the other two groups, $t(66) = 3.32$, $p < .001$, whereas there was no statistically significant difference between the phonological awareness and the control groups, $t(66) = 0.00$, $p = 1.00$. Note, however, that no statistically significant group differences were found when reading for the set of 10 novel words was analyzed, $F(2, 66) = 0.23$, $p = .80$. Hence, the immediate invented-spelling benefits seem limited to words for which children have practiced and thus may be developing some level of orthographic representation for. Given the low levels of pretested reading skills and the short intervention period, immediate transfer (of reading facilitation) to untrained words was not evident.

Facilitative effects: A learn-to-read task. The central hypothesis guiding this research was that invented spelling helps kindergarten children integrate phonological and orthographic information into initial lexical representations that can facilitate reading acquisition. To test this hypothesis directly, children were administered a learn-to-read task at posttest. Mean group performance on each of the five word-association recall trials is depicted in Figure 1. As evident in this presentation, the invented-spelling group appears to have outperformed the other groups on all trials. This was confirmed through a repeated measures analysis, which revealed a statistically significant between-group difference on the series of trials as a set, $F(2, 66) = 3.92$, $p < .05$, with a moderate effect size ($\eta^2 = .106$). There was a main effect of trial, $F(4, 264) = 82.56$, $p < .001$, and no significant interaction between group and trial, $F(8, 264) < 1$, $p = .65$. Planned orthogonal contrasts (Helmert) on the total scores over the series of trials confirmed the observed superior performance of the invented-spelling group over the two comparison groups (contrast estimate = 1.59, $SE = 0.59$, $p < .01$), whereas there was no statistically significant difference between the phonological awareness and the control groups (contrast estimate = 0.08, $SE = 0.62$, $p = .90$). These results provide direct support for the hypothesized role of invented spelling in learning to read words.

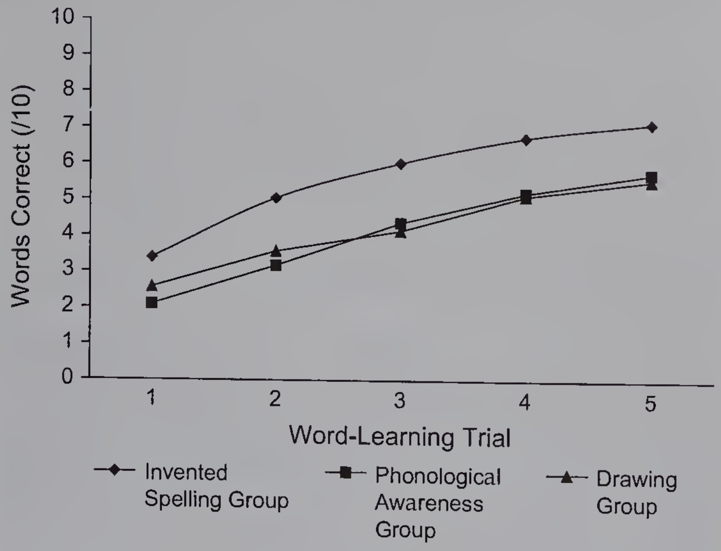


Figure 1. Mean number of words read on each trial of the learn-to-read task following completion of training.

Discussion

The findings of the present training study highlight the importance of invented spelling in the acquisition of early literacy skills and validate the training procedures used. Although all groups learned letter–sound knowledge, the invented-spelling group demonstrated more advanced invented spellings than did the other groups, and both the invented-spelling and the phonological awareness groups demonstrated superior phonological awareness. Invented-spelling training also benefited orthographic awareness and reading of words used in the training protocol.

Most importantly, the present results provide direct evidence for a causal role of invented spelling with developmentally appropriate feedback, in learning to read, thus confirming the hypotheses of Frith (1985) and Treiman (1993). In particular, the invented-spelling group learned to read more words in the learn-to-read task than did the children in the other groups. The present results offer direct training evidence that improvements in invented spelling can bring about an advantage in learning how to read. As previously discussed, invented spelling offers a powerful milieu in which to learn critical phonological awareness skills. Invented spelling, however, involves more than phonological awareness. It is important to note that in the present study, both the invented-spelling and the phonological awareness training brought about equivalent levels of posttested phonological awareness. Yet, the invented-spelling group was superior on the learn-to-read task. The present training results are thus interpreted as evidence that invented spelling coupled with feedback helps children hone and integrate a host of important skills that are potentially involved in learning how to read. In particular, invented spelling leads to increased phonological and orthographic awareness, and as proposed by Frith, invented spelling provides valuable insight and practice with the alphabetic code.

The exploratory nature of invented spelling encourages children to use an analytical approach that also promotes the integration of phonological and alphabetic information into initial lexical representations that connect phonological and orthographic information; as these lexical representations become refined, they may facilitate the acquisition of reading. This interpretation is congruent with the lexical quality hypothesis of Perfetti and Hart (2002) in that it highlights the importance of integrating information/representations from the domains of phonology and orthography. Invented spelling appears to provide children with an interactive activity to explore and analyze the written code and in the process make these proposed important associations between phonological and orthographic representations that will constitute the basis for complete lexical representations invoked in reading.

The invented-spelling group also outperformed the two control groups in reading the words used during training. Rieben et al. (2005) also reported superior reading for practiced words only for one of their invented-spelling training groups. This finding may be explained by the intensity of practice offered in the present study, along with the gradual developmental acquisition of lexical representations. In particular, the training involved spelling the same word more than once in a short time. This word-specific stimulation is more intensive than in typical child-directed spelling attempts, and this frequency of exposure may well be an important factor in learning to read words that were spelled. That is, with continued exposure and practice with the same word, the child is able to refine the respective lexical representation. This continuing refinement of representations is reflected in the described gradual, not stage-like, increases in spelling sophistication and eventual word reading skill seen within the invented-spelling training group.

In the present training study, the invented-spelling and phonological awareness groups had equivalent phonological awareness at posttest. Invented spelling with feedback, therefore, offers another context in which children learn valuable phonological awareness skills. This interpretation is consistent with Treiman (1993) and Mann et al. (1987), who proposed that invented spelling fosters phonological awareness growth, and is in accord with the recent training study reported by Martins and Silva (2006).

It should be noted that the invented-spelling training offered in the present study differed considerably from the combined phonological awareness and spelling training found in the phonological awareness intervention literature. That is, studies that combine word segmentation training with spelling (e.g., Uhry & Shepherd, 1993) explicitly teach phonological segmentation and then introduce letter tiles to directly combine phonological awareness and spelling. Children are then directly taught to spell words in the accepted conventional form. In the present invented-spelling training, however, there was no direct teaching of phonological segmentation and no explicit instruction bridging segmentation with conventional letter – sound mapping. Rather, children were encouraged to analyze and represent each word in their own way, followed by individualized feedback. The feedback was given to increase the phonological accuracy and complexity of children's spelling attempts. Specifically, the feedback respected the previously observed progression from initial nonalphabetic markings or random strings of letters to capturing the initial sounds of words, then marking the initial and final sounds, and then to the appearance of medial vowels in children's invented spellings (Gentry & Gillet, 1993). The progression also included the gradual elimination of extra letters. Martins and Silva (2006) successfully used similar feedback (i.e., provided models indicative of a higher phase of invented spelling) in their intervention research with Portuguese children. Although Portuguese is considered a much more transparent or shallow orthography than English, it too has inconsistencies in spelling (Defior, Martos, & Cary, 2002). In both the Martins and Silva training research and the present study, invented spelling along with developmentally appropriate feedback contributed to phonological awareness skills. These results, therefore, provide teachers with a valuable evidence-based alternative tool to promote phonological awareness. At the same time, future research should be conducted to evaluate the comparative benefits of different types of feedback, including the direct teaching of conventional spelling (see Rieben et al., 2005).

The invented-spelling training with feedback also increased children's implicit awareness of the orthographic rule in English that words must contain a vowel. Invented spelling can, therefore, offer a meaningful context from which children can become sensitive to orthographic rules and thus begin to store orthographic representations. These results are especially relevant, given recent contentions in the literature concerning when children learn orthographic information. Models such as the self-teaching theory of Share (1995, 1999) proposes that children must first decode to learn orthographic knowledge, whereas others have suggested that mere exposure to printed words may suffice (also see Caravolas, Kessler, Hulme, & Snowling, 2005; Wright & Ehri, 2007). The present findings suggest another route to the acquisition of orthographic knowledge, and that is through the exploration afforded by invented spelling.

Interestingly, the phonological awareness training group in the present research did not outperform the control group on measures of reading. This finding appears contradictory with the conclusions of the meta-analytic review by Ehri et al. (2001). Although the phonological awareness group in the present study did receive letter – sound training on the 13 sounds that composed the target words, the connection between the phonemes in the target words and their letter representations remained implicit. This type of implicit training has been used in previous research that has shown positive effects on reading (e.g., Byrne & Fielding-Barnsley, 1995), but it may be the case that in a short-term intervention like ours, the connection between phoneme segmentation or categorization and letter sounds must be made more explicit. For example, children in the classic study by Bradley and Bryant (1985) were taught with the help of plastic letters how common sounds across words were represented with letters. It remains to be tested, therefore, whether explicit instruction on using letters to segment words would be as beneficial to reading as would the invented-spelling intervention in the present study. We contend, however, that the knowledge afforded by invented spelling with feedback goes beyond that of explicit phonemic awareness training with letters. The work of Cunningham (1990) is helpful here. In her research on phonemic awareness training (without letters), she showed that training that highlighted the value and application of phonemic awareness produced greater effects than instruction that focused exclusively on the phonological awareness skills. In a similar vein, it is possible that invented spelling allows young kindergarten children to understand at a metacognitive level the function and value of letter-sound correspondences, which, in turn, further benefits their entry into reading. This possibility remains to be tested.

In the present 4-week training study, the invented-spelling group outperformed the comparison groups in learning how to decode words, reading familiar words, phonological awareness, and orthographic knowledge. Caution is warranted, however, because the magnitude of the between-group differences is modest. Still, the present short-term intervention does provide initial empirical support for the proposal originally brought forth by Chomsky (1971), and reiterated by Clarke (1988), that invented spelling can be incorporated into early literacy instruction. In particular, invented spelling combined with feedback that offers models just above the child's developmental level could be used by kindergarten teachers to promote growth in children's ability to capture spoken language in writing with their increasing knowledge of letters, as well as promote growth in phonological and orthographic awareness and, eventually, in learning to read words. It must be stressed that although the invented-spelling intervention has been described as *training* in the present study, the process is more accurately described as one of facilitation. That is, it must be stressed that by definition, invented spelling is a natural process (Burns & Richgels, 1989; Richgels, 1995). The intervention employed in the present study took advantage of each child's own spelling productions but then offered individualized feedback intended to guide the child to capture more completely the phonology of words, albeit in an unconventional manner. As noted previously, the feedback also focused the child's attention on unnecessary letters in their spellings. In addition to the inclusion of feedback, the present study worked with a relative small set of 13 letters, the words to be spelled were regular, and children were asked to spell each word four times. Teachers might consider using such strategies to optimize learning and facilitate their providing feedback to children in a classroom setting.

Implementation in kindergarten classrooms of invented spelling with developmentally appropriate feedback presupposes that teachers are familiar with the developmental trajectory of invented spelling. Whether teachers have this knowledge at present is uncertain. Nonetheless, teachers could acquire this knowledge as was shown with phonological awareness knowledge. Specifically, recent findings have shown that teachers who participated in workshops on how to implement explicit phonological awareness training in their kindergarten classrooms showed increased knowledge, changed their practices, and had children with greater literacy skills in Grade 1 compared to a control group of teachers and children (McCutchen et al., 2002). Most encouraging, the instructors in the present study reported informally that they found the developmental feedback intuitive and easy to implement.

With respect to further teaching applications, invented spelling had been criticized in the past as exposing children to erroneous orthography (Clarke, 1988). The feedback offered in the present study, however, and the general context of the training made it clear to the children that their spellings were not the same as an adult's. Invented spellings were, therefore, never reinforced as correct in a conventional sense. It should also be noted that Clarke (1988) found improved irregular word reading in Grade 1 classrooms where invented spelling was encouraged, and in the present research, invented-spelling training actually increased the children's orthographic awareness. Together then, these findings should go a long way in alleviating fears that encouraging invented spelling will impede acquisition of conventional spelling. Indeed, invented spelling offers a naturally occurring process that can be used as a springboard to facilitate phonological and orthographic awareness and, importantly, can help children along the pathway to learning how to read.

References

Ball, E., & Blachman, B. (1991). Does phoneme awareness training in kindergarten make a difference in early word recognition and developmental spelling? *Reading Research Quarterly, 26*, 49–66.

Bradley, L., & Bryant, P. E. (1985). *Rhyme and reason in reading and spelling*. Ann Arbor: University of Michigan Press.

Burns, J., & Richgels, D. (1989). An investigation of task requirements associated with invented spellings of 4-year-olds with above average intelligence. *Journal of Reading Behavior, 21*, 1–14.

Byrne, B., & Fielding-Barnsley, R. (1995). Evaluation of a program to teach phonemic awareness to young children: A 2- and 3-year follow-up and a new preschool trial. *Journal of Educational Psychology, 87*, 488–503.

Caravolas, M., Hulme, C., & Snowling, M. J. (2001). The foundations of spelling ability: Evidence from a 3-year longitudinal study. *Journal of Memory and Language, 45*, 751–774.

Caravolas, M., Kessler, B., Hulme, C., & Snowling, M. (2005). Effects of orthographic consistency, frequency, and letter knowledge on children's vowel spelling development. *Journal of Experimental Child Psychology, 92*, 307–321.

Cassar, M., & Treiman, R. (1997). The beginnings of orthographic knowledge: Children's knowledge of double letters in words. *Journal of Educational Psychology, 89*, 296–300.

Chomsky, C. (1971). Write first, read later. *Childhood Education, 47*, 296–300.

Clarke, L. K. (1988). Invented versus traditional spelling in first graders' writings: Effects on learning to spell and read. *Research in the Teaching of English, 22*, 281–309.

Clay, M. (1972). *Reading: The patterning of complex behavior*. London: Heinemann.

Cunningham, A. E. (1990). Explicit and implicit instruction in phonemic awareness. *Journal of Experimental Child Psychology, 50*, 429–444.

Defior, S., Martos, F., & Cary, L. (2002). Differences in reading acquisition development in two shallow orthographies: Portuguese and Spanish. *Applied Psycholinguistics, 23*, 135–148.

Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonological awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly, 36*, 250–287.

Ehri, L. C., & Wilce, L. S. (1987). Does learning to spell help beginners learn to read words? *Reading Research Quarterly, 20*, 47–65.

Elkonin, D. B. (1973). U.S.S.R. In I. Downing (Ed.), *Comparative reading: Cross national studies of behavior and processes in reading and writing* (pp. 551–580). New York: MacMillan.

Ferreiro, E. (1991). Psychological and epistemological problems on written representation of language. In M. Carretero, M. Pope, R. Simons, & J. Pozo (Eds.), *Learning and instruction: European research in an international context* (Vol. 3, pp. 157–173). New York: Pergamon Press.

Fletcher-Flinn, C. M., Shankweiler, D., & Frost, S. J. (2004). Coordination of reading and spelling in early literacy development: An examination of the discrepancy hypothesis. *Reading and Writing: An Interdisciplinary Journal, 17*, 617–644.

Fountas, I. C., & Pinnell, G. S. (1996). *Guided reading: Good first teaching for all children*. Portsmouth, NH: Heinemann.

Frith, U. (1985). Beneath the surface of developmental dyslexia. In K. Patterson, M. Coltheart, & J. Marshall (Eds.), *Surface dyslexia*. London: Lawrence Erlbaum.

Fry, E., Kress, J., & Fountoukidis, D. (2000). *The reading teacher's book of lists*. Paramus, NJ: Prentice Hall.

Gentry, J. R., & Gillet, J. W. (1993). *Teaching kids to spell*. Portsmouth, NH: Heinemann.

Henderson, E. H. (1981). *Learning to read and spell: A child's knowledge of words*. DeKalb, IL: N. Illinois University Press.

Levin, I., Both-De Vries, A., Aram, D., & Bus, A. (2005). Writing starts with own name writing: From scribbling to conventional spelling in Israeli and Dutch children. *Applied Psycholinguistics, 26*, 463–477.

Levin, I., & Korat, O. (1993). Sensitivity to phonological, morphological, and semantic cues in early reading and writing in Hebrew. *Merrill-Palmer Quarterly, 39*, 213–232.

Levy, B. A., Gong, Z., Hessels, S., Evans, M. A., & Jared, D. (2006). Understanding print: Early reading development and the contributions of home literacy experiences. *Journal of Experimental Child Psychology, 93*, 63–93.

Mann, V. A. (1993). Phoneme awareness and future reading ability. *Journal of Learning Disabilities, 26*, 259–269.

Mann, V. A., Tobin, P., & Wilson, R. (1987). Measuring phonological awareness through the invented spellings of kindergarten children. *Merrill-Palmer Quarterly, 33*, 365–391.

Martins, M. A., & Silva, C. (2006). The impact of invented spelling on phonemic awareness. *Learning-and-Instruction, 16*, 41–56.

McBride-Chang, C. (1998). The development of invented spelling. *Early Education and Development, 9*, 147–160.

McCutchen, D., Abbott, R. D., Green, L. B., Beretvas, S. N., Cox, S., Potter, N. S., et al. (2002). Beginning literacy: Links among teacher knowledge, teacher practice, and student learning. *Journal of Learning Disabilities, 35,* 69–86.

Morris, D., & Perney, J. (1984). Developmental spelling as a predictor of first grade reading achievement. *Elementary School Journal, 84,* 440–457.

O'Connor, R. E., & Jenkins, J. R. (1995). Improving the generalization of sound/symbol knowledge: Teaching spelling to kindergarten children with disabilities. *Journal of Special Education, 29,* 255–275.

Ouellette, G. P., & Sénéchal, M (2008). A window into early literacy: Exploring the cognitive and linguistic underpinnings of invented spelling. *Scientific Studies of Reading, 12,* 195–219.

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 189–213). Philadelphia: John Benjamins.

Read, C. (1971). Pre-school children's knowledge of English phonology. *Harvard Educational Review, 41,* 1–34.

Read, C. (1975). *Children's categorizations of speech sounds in English*. Urbana, IL: National Council of Teachers of English.

Richgels, D. J. (1995). Invented spelling ability and printed word learning in kindergarten. *Reading Research Quarterly, 30,* 96–109.

Rieben, L., Ntamakiliro, L., Gonthier, B., & Fayol, M. (2005). Effects of various early writing practices on reading and spelling. *Scientific Studies of Reading, 9,* 145–166.

Share, D. L. (1995). Phonological coding and self-teaching: Sine qua non of reading acquisition. *Cognition, 55,* 151–2218.

Share, D. L. (1999). Phonological recoding and orthographic learning: A direct test of the self-teaching hypothesis. *Journal of Experimental Child Psychology, 72,* 95–129.

Shatil, E., Share, D. L., & Levin, I. (2000). On the contribution of kindergarten writing to grade 1 literacy: A longitudinal study in Hebrew. *Applied Psycholinguistics, 21,* 1–21.

Tangel, D. M., & Blachman, B. A. (1992). Effect of phoneme awareness instruction on kindergarten children's invented spelling. *Journal of Reading Behavior, 24,* 233–261.

Tangel, D. M., & Blachman, B. A. (1995). Effect of phoneme awareness instruction on the invented spelling of first grade children: A one-year follow-up. *Journal of Reading Behavior, 27,* 153–185.

Treiman, R. (1993). *Beginning to spell.* New York: Oxford University Press.

Treiman, R. (1998). Why spelling? The benefits of incorporating spelling into beginning reading instruction. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 289–313). Mahwah, NJ: Lawrence Erlbaum.

Treiman, R., & Bourassa, D. C. (2000). Children's written and oral spelling. *Applied Psycholinguistics, 21,* 183–204.

Troia, G. (1999). Phonological awareness intervention research: A critical review of the experimental methodology. *Reading Research Quarterly, 34,* 28–52.

Uhry, J. K., & Shepherd, M. J. (1993). Segmentation/spelling instruction as part of a first grade reading program: Effects on several measures of reading. *Reading Research Quarterly, 28,* 218–233.

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive test of phonological processing*. Austin, TX: Pro-Ed.

Wright, D., & Ehri, L. 2007). Beginners remember orthography when they learn to read words: The case of doubled letters. *Applied Psycholinguistics, 28,* 115–133.

Appendix
Stimuli

Word reading assessment

Pretest
To
Come
Have
Are
Here
Lap[a]
Day[a]
No[a]
Ape[a]
Lady[a]

Posttest (same 10 words as mentioned previously, with 5 additional words added)
Lip
Bay
So
Ate
Bony

Invented-spelling assessment (pretest and posttest)

Lap[a]
Day[a]
No[a]
Ape[a]
Lady[a]
Sit
Rabbit
Pretty
Train
Boot

Orthographic awareness stimuli (posttest)

Vowels
GHW–GOW
GRLBR–GRIOR
CNPY–ENPY
PVGHTCN–PIGHTEN
KNDWL–KNOWL
SEAB–SCGL
CROSP–CRGSP
VO–VC
JACKEF–JDCKCF
NARBOW–NDRBCW

Double consonants
SALL–SSAL
PESS–PPES
NALL–NNAL
PPON–PONN
ROPP–RROP
SSEN–SENN
RRAS–RASS
NNEP–NEPP
TTAS–TASS
RETT–RRET

Learn-to-read task

Bed
Bee
Bend
Bone
Baby
Pin
Pie
Pond
Peel
Pony

Training stimuli for all groups and sequence of presentation

Ape
Eel
No
On
In
Day
See
Pie
Soap
Nail
Line
Pole
Pen
Bib
Lap
Sad
Rip
Rod
Lady
Ladder

---

[a] Items that overlap with the training stimuli.

Child Development, July/August 2008, Volume 79, Number 4, Pages 914–929

Coaching, Truth Induction, and Young Maltreated Children's False Allegations and False Denials

Thomas D. Lyon
University of Southern California

Lindsay C. Malloy and Jodi A. Quas
University of California, Irvine

Victoria A. Talwar
McGill University

This study examined the effects of coaching (encouragement and rehearsal of false reports) and truth induction (a child-friendly version of the oath or general reassurance about the consequences of disclosure) on 4- to 7-year-old maltreated children's reports ($N = 198$). Children were questioned using free recall, repeated yes – no questions, and highly suggestive suppositional questions. Coaching impaired children's accuracy. For free-recall and repeated yes – no questions, the oath exhibited some positive effects, but this effect diminished in the face of highly suggestive questions. Reassurance had few positive effects and no ill effects. Neither age nor understanding of the meaning and negative consequences of lying consistently predicted accuracy. The results support the utility of truth induction in enhancing the accuracy of child witnesses' reports.

The extent to which children's dishonesty may be affected by adult influences is of considerable theoretical and practical interest. Theoretically, knowledge concerning how adults influence children of various ages provides insight into cognitive and social development. Practically, children are routinely questioned about alleged experiences, both in day-to-day situations in which parents, teachers, and others inquire into mostly mundane events, and in more serious settings, such as legal contexts, where children's statements can have far-reaching implications.

The purpose of the present study was to examine factors influencing dishonesty in a large sample of maltreated children, with a specific focus on how adult influences affect dishonesty. Maltreated children are particularly likely to be subjected to pressures to disclose or conceal information (Malloy, Lyon, & Quas, 2007; Sas & Cunningham, 1995), and their honesty and dishonesty are often at issue in legal contexts (Brennan, 1994), thus highlighting the need to understand dishonesty in this population. Moreover, two forms of dishonest behavior should be considered: false denials of true events and false allegations of untrue events.

Several factors are potentially important in influencing children's dishonesty. First, many of children's lies concern actions involving others (Wilson, Smith, & Ross, 2003), and others often seek to influence children's honesty. The latter person may be called an "instigator," to reflect his or her potential influence on children's statements. Efforts by instigators to encourage and rehearse dishonesty in children will be referred to here as "coaching." Second, the person to whom children lie, referred to as the "recipient," may also exert some influence. In an applied context, the recipient is often an interviewer, although a recipient could be any person with whom children interact. Third, several characteristics in children likely affect their dishonesty, including their age and attitudes about the morality and utility of dishonesty.

Portions of this research were presented at the 17th annual meeting of the American Psychological Society, Los Angeles, CA, 2005, and the annual meeting of the American Psychology & Law Society, La Jolla, CA, 2005. Esi Benyarku, Alan Kang, Vanessa Negrete, David Reinert, Lauren Threatte, and Christina Tsai assisted in data collection and coding. We are grateful for the support of the presiding judge of the Los Angeles County Juvenile Court, the Los Angeles County Department of Children's and Family Services, Los Angeles County Children's Law Center, the Children's Services Division of Los Angeles County Counsel, and the Los Angeles County Child Advocate's Office. Preparation of this article was supported in part by National Institute of Child and Human Development Grant HD047290-01A2 and National Science Foundation Grant BCS-0241586.

Correspondence concerning this article should be addressed to Thomas D. Lyon, University of Southern California Law School, University of Southern California, 699 Exposition Boulevard, Los Angeles, CA 90089-0071. Electronic mail may be sent to tlyon@law.usc.edu.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7904-0008

Research relevant to each of these factors will be considered in turn.

Instigator Behavior

Research on the influence of third persons on children's dishonesty has usually examined situations in which the instigator asks the child to conceal the instigator's wrongdoing, often resulting in high rates of false denial (Bottoms, Goodman, Schwartz-Kenney, & Thomas, 2002; Bussey & Grimbeek, 1995; Ceci & Leichtman, 1992; Pipe & Wilson, 1994; Talwar, Lee, Bala, & Lindsay, 2004; Wilson & Pipe, 1989). Some research has also demonstrated that instigators can successfully encourage children to make false allegations, both in situations in which children's motivation was to sustain a game (Quas, Davis, Goodman, & Myers, 2007; Tate, Warren, & Hess, 1992) and in situations in which children's motivation was to conceal a parent's wrongdoing (Tye, Amato, Honts, Devitt, & Peters, 1999).

An open question, and one we sought to explore, is whether instigator influence may be overcome, particularly in situations where children have been extensively coached to make false reports. This is a typical concern in maltreatment cases; the defense often argues that abuse allegations are the product of coaching (Brennan, 1994), whereas the prosecution argues that inconsistencies and retractions are themselves the product of instigator influence (Summit, 1983).

Recipient Behavior

Two lines of research are relevant to how the behavior of a recipient affects children's dishonesty. One concerns the types of questions asked by the recipient, and the other concerns the statements made by the recipient to promote honesty. With regard to question type, research investigating children's concealment of transgressions finds that yes–no and other direct questions are more effective in reducing false denials than open-ended questions (Bottoms et al., 2002; Pipe & Wilson, 1994). However, the former types of questions can also increase the likelihood of false allegations (Thompson, Clarke-Stewart, & Lepore, 1997). Moreover, children coached to make false allegations are more likely to do so in response to yes–no or forced-choice questions than in free recall (Quas et al., 2007). Hence, more direct questions may exert their influence not by eliciting greater honesty but by increasing acquiescence.

The recipient may also make explicit requests for honesty, an approach we refer to as truth induction. Truth induction is premised on the notion that children's dishonest behavior is in part determined by their perceptions of the consequences of honesty and dishonesty (Bandura, 1991; Bussey, 1992). For example, in the context of abuse disclosure, abuse victims often report that their decision regarding whether and when to disclose was affected by their expectations about how others would react to their disclosure and the effects of disclosure on themselves and others close to them (Anderson, Martin, Mullen, Romans, & Herbison, 1993). Furthermore, a primary motive for young children's lies is to avoid punishment for misdeeds (Bussey, 1992; DePaulo, Jordan, Irvine, & Laser, 1982; Ekman, 1989; Stouthamer-Loeber & Loeber, 1986). At the same time, at least by 5 years of age, most children exhibit good understanding that lying is itself punished (Bussey, 1992; Lyon & Saywitz, 1999; Peterson, Peterson, & Seeto, 1983). Hence, even at a young age, children appear to evaluate the consequences of their disclosures, and their evaluations in turn affect their behavior.

Two approaches to truth induction have been researched. The first involves the recipient highlighting the consequences of honesty and dishonesty. Some research has found that discussion of the morality of lying increases children's accuracy (Huffman, Warren, & Larson, 1999; London & Nunez, 2002; but see Talwar, Lee, Bala, & Lindsay, 2002). Other research has made children's obligations to be honest more explicit by eliciting a promise to tell the truth, analogous to the administration of the oath in court. Talwar et al. (2002, 2004) found that eliciting a promise from 3- to 11-year-olds to tell the truth decreased their tendency to deny falsely that they had peeked at a toy or that their parent had broken a toy. In a procedure similar to that employed here, Lyon and Dorado (in press) found that eliciting a promise decreased 5- to 7-year-old maltreated children's tendency to conceal their own and an adult instigator's transgression (play that the instigator had warned "might get [them] in trouble"). Moreover, the positive effects of the promise could not be attributed to acquiescence, because promising did not increase children's tendency to claim falsely that they had played, even when asked leading tag questions (e.g., "You opened some of the doors, didn't you?").

A second approach to truth induction involves reassurance, in which the recipient reassures the child that the recipient will not punish or otherwise react negatively to the child's disclosure of wrongdoing. Reassurance addresses the fact that, in some contexts, children may anticipate that honesty (rather than dishonesty) will be punished. To date, only one study has examined the potential for reassurance to decrease dishonesty. Lyon and Dorado (in press) provided maltreated children reassurance about disclosing forbidden play. The reassurance was specific in that the recipient explicitly mentioned the target play (i.e., "It's okay if you played with the toy house"). Specific reassurance decreased children's tendency to make false denials but increased false alarms among those children who exhibited some difficulty in understanding the meaning and negative consequences of lying. Hence, although some benefits of specific reassurance may exist, such reassurances also risk increasing acquiescence and false alarms (cf. Goodman, Batterman-Faunce, Schaaf, & Kenney, 2002).

No research has assessed whether reassurance decreases dishonesty without increasing false alarms if it is less specific, that is, if the recipient reassures the child without naming the wrongdoing in question.

In this study, we sought to examine the effects of eliciting a promise to tell the truth or general reassurance on children (referred to jointly as truth induction) while varying the question type from less to more leading (free recall, yes – no, and suppositional). In doing so, we sought to identify the optimal means to reduce dishonesty. In order to distinguish between honesty and acquiescence, we examined situations in which children were motivated to either falsely claim or falsely deny behavior.

Child Characteristics

Children's age and beliefs about the meaning and morality of dishonesty are potentially important in predicting dishonesty. First, the tendency to make false denials regarding transgressions emerges early in the preschool years and develops rapidly during this period, with substantial percentages of children as young as 3 years of age denying their own transgressions (Bussey, Lee, & Grimbeek, 1993; Lewis, Stanger, & Sullivan, 1989; Polak & Harris, 1999; Talwar, Gordon, & Lee, 2007; Talwar & Lee, 2002; Talwar et al., 2002). Similar age effects have been found regarding children's denial of transgressions of instigators who discouraged disclosure (Bottoms et al., 2002; Bussey & Grimbeek, 1995; Ceci & Leichtman, 1992). Relative to older children, 3-year-olds have a limited understanding of the effects of their statements on others' beliefs and are less aware of the negative consequences of revealing transgressions (Bottoms et al., 2002; Bussey & Grimbeek, 1995; Polak & Harris, 1999; Talwar & Lee, 2002; Tate et al., 1992). Moreover, younger children may find it more difficult to maintain a false story in light of the cognitive demands of monitoring one's responses for consistency and plausibility (Polak & Harris, 1999; Talwar & Lee, 2002; Talwar et al., 2007; Tate et al., 1992). Maltreated children, who tend to suffer from delays in development (Lyon & Saywitz, 1999; Shonk & Cicchetti, 2001), would likely exhibit similar patterns, albeit at somewhat later ages.

Second, children's understanding of the meaning and consequences of lying also develops during the preschool years (Bussey, 1992; Lyon & Saywitz, 1999; Peterson et al., 1983) and may have implications for their dishonesty. Although most research to date has not found an association between understanding and dishonesty (Feben, 1985; Goodman, Aman, & Hirschman, 1987; London & Nunez, 2002; Pipe & Wilson, 1994; Talwar et al., 2002), there are several reasons why a relation may have been missed. First, some research focused on memory errors rather than deliberately false statements (Feben, 1985; Goodman et al., 1987). Memory errors are not lies and therefore would be minimally influenced by a child's attitudes regarding honesty. Second, several studies examining the relation between children's understanding of lying and their tendency to deny a transgression had limited power to detect significant relations, either because of minimal variability in understanding (London & Nunez, 2002) or because understanding was measured with only a few questions (Pipe & Wilson, 1994; Talwar et al., 2002). Third, from a legal perspective, understanding of the meaning and morality of lying does not itself predict honesty but establishes an understanding of the oath. The supposition is that those who understand the oath are more likely to be influenced by the oath. Indeed, Talwar et al. (2004) found a positive relation between understanding of lying and honesty among children who had promised to tell the truth. Fourth, virtually all the research has examined possible relations between understanding and false denial of transgressions, without considering deliberately false allegations. Lyon and Dorado (in press) found that maltreated children who exhibited good understanding of the meaning and morality of the oath were less susceptible to making false allegations in response to leading tag questions than children who failed to exhibit good understanding.

Because of the potential effects of age on children's responsiveness to coaching and truth induction, we evaluated children across an age range during which one would expect developmental change in their abilities and proclivities to behave dishonestly. We also assessed children's attitudes about honesty and dishonesty in relation to coaching and truth induction to determine whether children's understanding predicted their behavior.

Present Study

In the present study, we examined the effects of instigator and recipient behavior on 4- to 7-year-old maltreated children's false allegations and false denials. During an initial phase, an instigator and a child either played or did not play with a toy house. Afterward, for half of the children, the instigator claimed that their behavior violated the recipient's expectations and provided extensive coaching so that the child would make a false report. That is, children who played were coached to deny that they had played, and children who did not play were coached to claim that they did. A recipient then interviewed the child about what happened and made increasingly strong suggestions that play with the house had occurred. The interview included free-recall, yes – no, and suppositional questions, which presupposed that play had occurred. During each phase of questioning, the recipient either gave no instructions (the control condition) or used one of the two types of truth induction. In the oath condition, children were asked to promise to tell the truth about what had occurred, and in the reassurance condition, children were reassured that disclosure of wrongdoing would not lead to punishment. The recipient also administered a task to assess children's understanding of the meaning of truth and lie and the negative consequences of lying. All procedures were approved by the researchers' institutional review boards as well as the presiding judge of the juvenile court and the agencies who work with maltreated children.

Based on prior research, three sets of hypotheses were tested. First, with respect to coaching, we predicted that dishonesty would be higher among coached children than among noncoached children, with respect to both children coached to make false denials and children coached to make false allegations. Second, with respect to truth induction, we predicted that the oath and reassurance would decrease dishonesty. We expected that reassurance would be particularly effective with respect to children who had played because of the likelihood that children would perceive such play as a transgression. However, reassurance should not increase false allegations among children who had not played because the form of reassurance used in this study did not explicitly mention play (in contrast to findings that specific reassurance increased false allegations among some children; Lyon & Dorado, in press). We further hypothesized that the efficacy of truth induction would depend upon the types of questions asked by the recipient, with induction least effective when coupled with the highly leading suppositional questions that presuppose play had occurred. Third, with respect to individual differences, we predicted that older children would be both more likely to lie and more likely influenced by truth induction than younger children and that greater understanding of the meaning and consequences of lying would be associated with more honest responding, at least in the truth induction conditions.

Method

Participants

Participants included one hundred and ninety-eight 4- to 7-year-old children (99 boys and 99 girls) who were awaiting a court appearance in the Los Angeles County Dependency Court. All children had been removed from the custody of their parent or guardian due to substantiated maltreatment. Children were ineligible to participate if they were Spanish speaking (either officially recognized as Spanish speaking by the court or clearly incapable of communicating with the researcher in English) or if they were awaiting an adjudication hearing on the day of their appearance in court (at which they might have to testify). The final sample included ninety-eight 4- to 5-year-olds (M = 63 months, range = 48 – 71 months) and one hundred 6- to 7-year-olds (M = 83 months, range = 72 – 95 months). Their ethnicity was diverse: Forty-four percent were Latino, 40% were African American, 10% were non-Hispanic Caucasian, and 6% were other or unknown.

Materials

Oath-taking competency task (Lyon & Saywitz, 1999). This task consists of questions about eight pictures, four regarding children's understanding of the meaning of "truth" and "lie" and four regarding children's understanding of the negative consequences of lying. For the meaning questions, pictures of two child characters looking at a single object are presented. One character labels the object correctly, and the other character labels the object incorrectly, with the labels depicted as pictures within talk bubbles. Children are asked which character "told the truth" (two trials) or "told a lie" (two trials), thereby assessing their understanding that truth refers to accurate statements and lie refers to inaccurate statements. For the consequence questions, pictures of two child characters talking to an adult (judge, doctor, social worker, or grandmother) are presented. One character is described as telling "the truth," and the other is described as telling "a lie." Children point to the character who "will get in trouble," with correct responses demonstrating children's understanding that lying has negative consequences. Correct responding to the meaning and consequence questions would qualify a child witness as competent to take the oath (and thus to testify; Lyon, 2000).

Toy house. A toy house was constructed out of Lego™ building blocks. The house had several small doors on each side, with small toys hidden behind each door. The house was placed on a revolving platform so that each side could be easily accessed.

Procedure

For ease of exposition, the recipient will be referred to as "she" and the instigator as "he." All recipients were female, and two of the four instigators were male; instigator–recipient pairs were equally distributed across the conditions. The recipient invited eligible children to participate in the study and obtained their assent. She then administered the oath-taking competency task and afterward told the child that she had to retrieve a forgotten form in her office. She asked the child to wait for her, and to help the child wait, she set a timer to 5 min and told the child that she would be back before the timer went off. Upon leaving, she gestured to the toy house on the floor and said that they would play with it when she returned.

Instigator–child interaction. Shortly after the recipient left, the instigator entered the room and introduced himself. He engaged the child in a guessing game with a coin, in which one person guessed in which fist the other person held a coin. During some of the trials, the instigator would hide the coin so that neither fist held the coin and remark, "It's fun to trick people." Next, the instigator engaged the child in one of the four conditions: play, play coach, no-play, or no-play coach. The instigator spent approximately 5 min with each child, leaving shortly before the timer rang.

In the *play* ($n = 52$) and *play coach* conditions ($n = 49$), after playing the guessing game, the instigator noticed the toy house, lifted it onto the table, and engaged the child in play with the toy. He ensured that both he and the child opened doors, looked at the toys inside, and took the toys out. He labeled his actions, explicitly encouraged the child to engage in each action, and labeled the child's actions, thereby making the interaction with the house and toys quite salient to the child.

In the play condition, the instigator continued to play with the child and the toy until it was time to return the toy to the floor. He then said good-bye and left. In the play coach condition, the instigator interrupted the play to coach the child to deny play. He said, "We might get in trouble if anyone finds out that we played with the toy. I need you to do something for me. When that lady comes back, she is going to ask you what we did. Don't tell her we played with the toy. She is going to ask you a bunch of questions about what we did in here. Tell her we played with the coin, but don't tell her about the house. I'll come back in a little while and tell her what happened. Can you help me, and trick her so she doesn't know we played with the toy?" Children who expressed reluctance were given additional encouragement to help. The instigator then practiced asking the child a few questions (both free recall and yes–no) to ensure that the child understood the instructions (e.g., "What did you do while I was gone?"). If the child disclosed that he or she played the guessing game in response to a practice question, the instigator responded with positive feedback. If the child reported playing with the house, the instigator instructed, "Try saying we played with the coin. Don't tell her we played with the house yet." Practice questions were delivered twice if the child gave the coached responses and thrice if the child did not follow the lead. As the instigator prepared to leave, he told the child, "Remember—tell her that we played with the coin, but please don't tell her about the toy. I'll tell her what happened later."

Children in the two no-play conditions did not engage in play with the toy house. In the *no-play* condition ($n = 49$), the instigator played the guessing game with the child the entire time he was in the room and never mentioned the toy house. In the *no-play coach* condition ($n = 48$), after playing the guessing game the instigator said, "Oh no! You know what, we were supposed to play with that toy on the floor! I need you to do something for me. When that lady comes back, she is going to ask you what we did. Let's tell her that we played with the toy." The instigator then followed a script that paralleled the play coach condition, except that he coached the child to claim that they had played with the toy house. The instigator explained that the toy house had doors with toys inside and practiced questions in which the child would falsely claim that they had touched the toy house, opened the doors, and taken toys out.

Recipient–child interview. When the recipient returned, she immediately asked a preinduction free-recall question, "What happened in here while I was gone?" followed by two "tell-me-more" prompts ("You said x. Tell me more about x.") Pilot testing revealed that children often blurted out their coached stories upon the recipient's arrival, making it impossible to deliver truth induction instructions. Therefore, we included this free-recall question, which enabled us to examine the influence of instigator coaching on children's reports prior to any truth induction.

Next, the recipient delivered one of the three truth induction instructions. In the *oath* condition ($n = 67$), the recipient said, "It's very important that you tell me the truth. Can you promise that you will tell me the truth? Will you tell me any lies?" (The latter question, which always elicited a "no" response, was asked to ensure that children were not simply acquiescing when answering "yes" to the first question.) In the *reassurance* condition ($n = 66$), the recipient said, "I talk to kids all the time about things that make them feel bad. If something happened that you feel bad or sad about, it's totally o.k. You won't get in any trouble with me at all. We can try to fix it and make it better." In the *control* condition ($n = 65$), the recipient gave no instructions. The recipient then asked the postinduction free-recall question ("Now, what happened in here while I was gone?") followed by two tell-me-more prompts.

The recipient then told the child, "I'm going to ask some more questions to make sure I have this right." Children in the oath condition were reminded, "Remember, you promised you would tell me the truth and that you won't tell me any lies." Children in the reassurance condition were reminded, "Remember you won't get in trouble with me for telling what happened."

The recipient then asked the repeated yes–no questions. The first two oriented the child to the instigator's activities: "While I was gone, did anyone come in the room?" and, if the child answered "yes," "Was it a man or a lady?" (the interviewer then used the appropriate pronoun for the remaining questions). If the child denied that anyone had come in the room, the interviewer omitted all subsequent yes–no questions about the instigator (and the child was scored as answering "no" to these questions). There were six yes–no questions about the child's and instigator's interaction with the toy house, with the order of asking about the instigator's and child's interaction counterbalanced. The questions were "While I was gone, did you [he] touch any of the doors?" "Did you [he] open any of the doors?" and "Did you [he] take any of the toys out?" Each question was immediately repeated in a skeptical tone.

The recipient then insisted that the child and instigator had played with the toy house, saying that he "comes in all the time and plays with my toys. I *know* that he came in and that *both* of you played with my toy house." The recipient repeated the truth induction instructions for children in the oath and reassurance conditions and then asked the six suppositional questions. These questions presupposed that the child and instigator had played with the house, and the order of the questions about the child and instigator was counterbalanced. The questions were "When you [the man] touched the doors, how many did you [he] touch?" "When you [the man] opened the doors, were you [was he] happy or mad?" and "When you [the man] took out the toys, did you [the man] play with them or look at them?"

To recap, there were four instigator conditions, two in which the child played with the toy house (play and play coach) and two in which the child did not (no-play and no-play coach). For each play or no-play condition, there was a corresponding condition in which the instigator coached the child to provide a false report. There were three truth induction conditions: control (no instructions), oath, and reassurance. The instigator and recipient conditions were fully crossed so that there were 12 groups and at least 16 children in each combination of instigator and recipient condition. All children were asked three types of questions: free recall, repeated yes–no, and suppositional.

Debriefing. The instigator then entered the room and fully disclosed what had happened while the recipient was away. The child was fully debriefed. Specifically, the recipient told the child that the instigator was her friend, that the 2 worked together at school, and that she knew the instigator would play with the child. She explained that the purpose of the study was to find the ways of helping children to tell the truth. She acknowledged that it is sometimes difficult to tell what happened but emphasized that it is important to tell the truth and thanked the child for helping.

Results

Children's interview responses were scored for three types of information: house and nonhouse free-recall details, accuracy to yes–no questions, and acquiescence to suppositional questions. For all scores, responses from 15% of the children (randomly selected across study conditions) were coded by two independent raters. Proportion agreement for free recall was 90%, and kappas for the other variables were $\geq .92$. Discrepancies were resolved through discussion. Preliminary analyses revealed no gender differences in performance or order (oath-taking competency task meaning stories first or consequence stories first, questions about instigator behavior first or child behavior first) effects. These factors are not considered further.

Free Recall

Children were asked recall questions twice, once before any truth induction and once after. Free-recall scores included the number of house and nonhouse details. House details were coded liberally and included any reference to the house that might be interpreted as play with the house (e.g., "I was playing with that [points to house on floor]"; "There are toys inside those [referring to windows on house]"). Nonhouse details referred to any other factual information provided by the child about their behavior while the recipient was absent (e.g., "We was playing the coin game"; "There was a lady"; "She had a penny"). Children's mean house and nonhouse details provided preinduction and postinduction are presented in Tables 1 and 2, respectively.

House details were correct for children who actually played with the house (i.e., children in the play and play coach conditions) but incorrect for children who did not play with the house (i.e., children in the no-play and no-play coach conditions). Depending on whether children had in fact played, both coaching and truth induction were expected to have opposite effects: Children who played would say less about play when coached to deny such play and more under truth induction, whereas children who did not play would say more about play when coached to do so and less under truth induction. Therefore, we assessed children's free recall separately for children who played and who did not play.

Table 1
Means (Standard Deviations) for Preinduction Free-Recall House and Nonhouse Details

	House	Nonhouse
Play (n = 52)		
4- to 5-year-olds	1.33 (1.55)	1.58 (1.28)
6- to 7-year-olds	1.93 (1.65)	2.46 (1.29)
M	1.65 (1.62)	2.06 (1.35)
Play coach (n = 49)		
4- to 5-year-olds	0.92 (1.82)	1.72 (1.31)
6- to 7-year-olds	0.54 (1.47)	1.67 (1.20)
M	0.74 (1.66)	1.69 (1.25)
No-play (n = 49)		
4- to 5-year-olds	0.04 (0.20)	2.38 (1.88)
6- to 7-year-olds	0.20 (0.71)	2.88 (1.09)
M	0.12 (0.53)	2.63 (1.54)
No-play coach (n = 46)		
4- to 5-year-olds	2.65 (2.35)	1.39 (1.31)
6- to 7-year-olds	4.26 (2.05)	1.04 (1.15)
M	3.46 (2.33)	1.22 (1.23)

Children's first recall (preinduction) performance reflected the influences of coaching and age before any truth induction. We conducted analyses of variance (ANOVAs) on the number of house and nonhouse details with age and instigator condition as between-subjects factors and type of detail (house vs. nonhouse) as a within-subjects factor. For children who played, significant main effects of instigator (play vs. play coach), $F(1, 97) = 9.93$, $p < .01$, partial $\eta^2 = .093$, and type of detail, $F(1, 97) = 9.74$, $p < .01$, partial $\eta^2 = .091$, emerged, as did a significant interaction between instigator and age, $F(1, 97) = 5.96$, $p = .01$, partial $\eta^2 = .058$. Children provided more nonhouse details (M = 1.88), than house details (M = 1.21). Also, although children provided fewer details in the play coach condition than in the play condition, this was qualified by the interaction with age. Older but not younger children produced fewer details when they were coached to deny play than when not coached, $t(97) = 2.85$, $p < .01$, $d = .98$, and among the children who played, older children provided more details than younger children, $t(97) = 1.85$, $p < .05$, $d = .67$. Summed house and nonhouse details are as follows: older play $M = 4.39$, play coach $M = 2.21$, younger play $M = 2.92$, and play coach $M = 2.64$.

The effects of coaching can also be examined by considering the number of children in each group who mentioned any play with the house. Because of the age differences for children in the play conditions, we analyzed the age groups separately. Although the younger children were only marginally less likely to disclose play when coached not to do so, 32% (8 of 25) in the play coach condition versus 58% (14 of 24) in the play condition, $\chi^2(3) = 3.43$, $p = .06$, $\varphi = .26$, the older children were clearly influenced by coaching: 27% (4 of 24) disclosed play in the play coach condition compared to 75% (21 of 28) in the play condition, $\chi^2(3) = 17.62$, $p < .001$, $\varphi = .58$.

Next, we examined preinduction performance among children who had not played with the house and hence for whom house details were false. If children in the no-play coach condition revealed the fact that the instigator had coached them to claim that they had played with the house (n = 2), they were coded as having provided zero house details because they were not reporting that they had actually played with the house. An ANOVA on the number of house and nonhouse details with age and instigator as between-subjects factors and type of detail as a within-subjects factor revealed significant main effects of age, $F(1, 93) = 5.98$, $p < .05$, partial $\eta^2 = .06$, and instigator, $F(1, 93) = 21.03$, $p < .001$, partial $\eta^2 = .19$,

Table 2

Means (Standard Deviations) for Postinduction Free-Recall House and Nonhouse Details

	Control (n = 64)		Oath (n = 66)		Reassurance (n = 66)	
	House	Nonhouse	House	Nonhouse	House	Nonhouse
Play (n = 52)						
4- to 5-year-olds	0.50 (1.07)	0.63 (0.92)	2.00 (2.39)	1.88 (1.96)	1.38 (1.19)	2.00 (2.00)
6- to 7-year-olds	2.00 (2.56)	1.75 (1.04)	1.60 (1.65)	2.60 (1.84)	1.70 (2.41)	0.90 (1.29)
M	1.25 (2.05)	1.19 (1.11)	1.78 (1.96)	2.28 (1.87)	1.56 (1.92)	1.39 (1.69)
Play coach (n = 49)						
4- to 5-year-olds	0.38 (1.06)	0.38 (0.52)	1.67 (2.06)	0.89 (1.05)	1.25 (2.38)	0.88 (0.64)
6- to 7-year-olds	0.63 (1.19)	1.13 (1.36)	0.88 (1.46)	2.00 (1.51)	0.00 (0.00)	0.88 (0.64)
M	0.50 (1.10)	0.75 (1.06)	1.29 (1.79)	1.41 (1.37)	0.63 (1.75)	0.88 (0.62)
No-play (n = 49)						
4- to 5-year-olds	0.00 (0.00)	2.13 (1.81)	0.13 (0.35)	1.50 (1.20)	0.25 (0.71)	1.13 (1.25)
6- to 7-year-olds	0.11 (0.33)	1.67 (1.80)	0.00 (0.00)	1.63 (1.69)	0.00 (0.00)	2.50 (1.77)
M	0.06 (0.24)	1.88 (1.76)	0.06 (0.25)	1.56 (1.41)	0.13 (0.50)	1.81 (1.64)
No-play coach (n = 46)						
4- to 5-year-olds	2.25 (2.92)	0.13 (0.35)	0.43 (0.79)	1.29 (1.11)	1.25 (2.38)	1.13 (1.36)
6- to 7-year-olds	1.29 (1.38)	0.43 (0.79)	1.13 (2.23)	1.13 (0.99)	1.75 (2.38)	1.00 (1.07)
M	1.80 (2.31)	0.27 (0.59)	0.80 (1.70)	1.20 (1.01)	1.50 (2.31)	1.06 (1.18)

as well as a significant interaction between instigator and type of detail, $F(1, 93) = 92.75$, $p < .001$, partial $\eta^2 = .50$, and a three-way interaction among age, instigator, and type of detail, $F(1, 93) = 5.23$, $p < .05$, partial $\eta^2 = .05$. Older children ($M = 4.10$) provided more details overall than younger children ($M = 3.17$), and children coached to falsely claim play ($M = 4.54$) provided more details overall than children not coached ($M = 2.74$). The interaction between instigator and type of detail was attributable to the fact that coaching children to claim house play led to an increase in the number of house details (no-play coach $M = 3.31$ and no-play $M = 0.12$) and a decrease in the number of nonhouse details (no-play coach $M = 1.23$ and no-play $M = 2.63$), $ts(93) > 4.18$, $p < .001$, $ds > .90$. The three-way interaction was attributable to the fact that the increase in the number of house details relative to the number of nonhouse details in the no-play coach condition was larger among the older children.

Whereas both age groups in the no-play coach condition provided a comparable number of nonhouse details, older children provided a substantially larger amount of house details than younger children, $t(46) = 2.61$, $p < .05$, $d = .73$ (Table 1).

We were able to assess the effects of coaching among children who did not play by counting the number of children who mentioned play in their preinduction free recall. Coaching clearly increased their tendency to claim play with the house; 6% (3 of 49) did so in the no-play condition compared to 82% (39 of 48) in the no-play coach condition, $\chi^2(3) = 55.74$, $p < .001$, $\varphi = .75$.

Children's second recall reflected possible influences of truth induction in addition to coaching and age (Table 2). First, children who played with the house were considered. We conducted an ANOVA on the number of house and nonhouse details with age, instigator, and induction as between-subjects factors and type of detail as a within-subjects factor. A significant main effect of instigator, $F(1, 89) = 8.41$, $p < .001$, partial $\eta^2 = .09$, revealed that children in the play condition provided more details than children in the play coach condition ($Ms = 3.14$ and 1.82). A significant main effect of induction, $F(2, 89) = 4.03$, $p < .05$, partial $\eta^2 = .08$, showed that children provided more information when administered the oath ($M = 3.36$) than when provided no instructions ($M = 1.84$), $t(89) = 1.93$, $p < .05$, $d = .49$, and, to some extent, than when provided reassurances ($M = 2.24$), $t(89) = 1.46$, $p < .10$, $d = .44$. However, this main effect was qualified by a significant interaction between age and induction, $F(2, 89) = 3.09$, $p = .05$, partial $\eta^2 = .07$. The induction effects were only evident among the 4- and 5-year-olds (oath $M = 3.22$, control $M = 0.94$, reassurance $M = 2.75$), $F(2, 46) = 4.59$, $p < .05$, partial $\eta^2 = .17$: Younger children who received the oath provided more information than children who received control instructions, $t(46) = 2.03$, $p < .05$, $d = .81$, as did younger children who received reassurance, $t(46) = 1.63$, $p = .05$, $d = .71$. The 6- and

7-year-olds did not differ in the amount of information provided based on induction (oath $M = 3.54$, control $M = 2.75$, reassurance $M = 1.74$). With respect to children who did not play with the toy but were coached to say that they had (the no-play coach condition), children were excluded if they explicitly disclosed the trick preinduction ($n = 2$) and children who disclosed the trick postinduction had house detail scores set to zero ($n = 5$; 4 in the oath condition and 1 in the reassurance condition). An ANOVA on the number of house and nonhouse details with age, instigator, and induction as between-subjects factors and type of detail as a within-subjects factor yielded a significant effect due to type of detail, $F(1, 83) = 7.49$, $p < .01$, partial $\eta^2 = .083$, and an interaction between instigator and detail, $F(1, 83) = 25.63$, $p < .001$, partial $\eta^2 = .236$, such that coaching led children to provide an equal number of house and nonhouse details, whereas children who were not coached to make false claims provided more nonhouse details, $t(83) = 5.59$, $p < .001$, $d = 1.73$—no-play coach: house $M = 1.35$, nonhouse $M = 0.85$; no-play: house $M = 0.08$, nonhouse $M = 1.76$.

The effects of induction can also be assessed by calculating the proportion of children who were initially dishonest but behaved honestly postinduction. With respect to children who played with the toy, examination of the counts suggested that the oath had an effect on children coached to conceal house play, whereas reassurance had an effect on children who failed to disclose play without any coaching. That is, for children in the play coach group, 5 of 12 children (42%) in the oath group who initially said nothing about play revealed play postinduction, compared to 3 of 25 children (12%) in the control and reassurance groups, $\chi^2(3) = 6.93, p < .01, \varphi = .43$. For children in the play group, 7 of 8 children (88%) in the reassurance group who had not initially mentioned play did so postinduction, compared to 4 of 9 children (44%) in the control and oath groups, $\chi^2(3) = 3.44, p = .06, \varphi = .44$.

Summary. As expected, coaching reliably affected children's tendency to disclose play with the toy house, both when they were coached to deny play that actually occurred and when coached to allege play that had not occurred. Older children were more affected by coaching than younger children. The oath and reassurance increased younger children's willingness to acknowledge truthfully that they had played with the toy house. For children who had not played but had been coached to make false allegations, truth induction had no reliable effects on disclosure. Thus, truth induction had some positive effects on increasing true disclosures among the younger children and no negative effects on false allegations.

Repeated Yes–No Questions

Following the free-recall prompts, the interviewer reminded children of the appropriate truth induction instructions and asked a series of repeated yes–no questions asking about play with the toy house. Accuracy proportion scores were created separately for questions about the child's and the instigator's behavior. For children who played with the house, accuracy reflected true disclosures, and for children who had not played with the house, accuracy reflected true denials. Less than 1% of children's responses were "do not know" or unscorable (e.g., unintelligible), and these responses were not considered further. Analyses comparing children's response accuracy between each initial question and when it was repeated revealed no significant differences (both directly or in conjunction with children's age, instigator, or induction condition), all $Fs(1 \text{ or } 2, 174) < 1.65$, *ns*. In fact, very few children changed any of their responses across the repeated questions ($M = 0.04$). Hence, children's responses to the initial and repeated questions were combined to create overall accuracy scores. Means are presented in Table 3.

We conducted an ANOVA on the children's accuracy scores with age, instigator (play, play coach, no-play, and no-play coach), and induction as between-subjects factors and actor (child vs. instigator) as a within-subjects factor. A significant main effect of instigator, $F(3, 174) = 36.84$, $p < .001$, partial $\eta^2 = .39$, reflected the fact that children in the no-play condition ($M = 0.93$) were considerably more accurate in denying house play than all other children, including children who were coached to claim falsely that they had played ($M = 0.50$). Children in the play condition ($M = 0.60$) were more accurate in acknowledging play than children coached to deny play falsely ($M = 0.28$), $ts > 4.80$, $ps < .01$, $ds > .98$.

A significant main effect of induction, $F(2, 174) = 11.85$, $p < .001$, partial $\eta^2 = .12$, showed that children who were administered the oath answered a significantly greater proportion of questions correctly ($M = 0.72$), than did children administered no induction instructions ($M = 0.50$), or children provided reassurance ($M = 0.51$), $ts > 3.20$, $ps < .001$, $ds > .56$.

There was also a significant three-way interaction among actor, age, and instigator, $F(3, 174) = 3.28$, $p < .05$, partial $\eta^2 = .05$, which revealed that in the play condition (but in none of the other instigator conditions), age interacted with instigator, $F(1, 46) = 5.37$, $p < .05$, partial $\eta^2 = .10$. Examination of the means

Table 3
Means (Standard Deviations) for Proportion Accuracy Scores to Repeated Yes–No Questions

	Control (n = 65)			Oath (n = 67)			Reassurance (n = 66)		
	Child	Instigator	*M (SD)*	Child	Instigator	*M (SD)*	Child	Instigator	*M (SD)*
Play (n = 52)									
4- to 5-year-olds	0.54 (0.48)	0.56 (0.44)	0.55 (0.30)	0.58 (0.50)	0.67 (0.44)	0.63 (0.42)	0.27 (0.40)	0.79 (0.36)	0.53 (0.24)
6- to 7-year-olds	0.35 (0.38)	0.29 (0.39)	0.32 (0.34)	0.88 (0.25)	0.68 (0.39)	0.78 (0.23)	0.78 (0.42)	0.67 (0.40)	0.73 (0.32)
M	0.45 (0.43)	0.43 (0.43)	0.44 (0.33)	0.75 (0.40)	0.68 (0.40)	0.71 (0.33)	0.56 (0.47)	0.72 (0.38)	0.64 (0.29)
Play coach (n = 49)									
4- to 5-year-olds	0.33 (0.47)	0.23 (0.43)	0.28 (0.41)	0.38 (0.48)	0.26 (0.43)	0.32 (0.27)	0.29 (0.45)	0.29 (0.42)	0.29 (0.33)
6- to 7-year-olds	0.25 (0.38)	0.06 (0.18)	0.16 (0.33)	0.48 (0.43)	0.63 (0.45)	0.55 (0.37)	0.04 (0.12)	0.13 (0.35)	0.08 (0.24)
M	0.29 (0.42)	0.15 (0.33)	0.22 (0.32)	0.43 (0.45)	0.43 (0.47)	0.43 (0.33)	0.17 (0.34)	0.21 (0.39)	0.19 (0.30)
No-play (n = 49)									
4- to 5-year-olds	0.98 (0.06)	0.92 (0.24)	0.95 (0.12)	0.98 (0.06)	1.00 (0.00)	0.99 (0.03)	1.00 (0.00)	0.64 (0.44)	0.84 (0.22)
6- to 7-year-olds	0.89 (0.33)	0.86 (0.29)	0.87 (0.29)	0.98 (0.06)	1.00 (0.00)	0.99 (0.03)	0.88 (0.35)	1.00 (0.00)	0.94 (0.18)
M	0.93 (0.24)	0.88 (0.26)	0.91 (0.22)	0.98 (0.06)	1.00 (0.00)	0.99 (0.03)	0.94 (0.25)	0.84 (0.34)	0.89 (0.20)
No-play coach (n = 48)									
4- to 5-year-olds	0.38 (0.52)	0.50 (0.53)	0.44 (0.50)	0.90 (0.23)	0.75 (0.46)	0.82 (0.32)	0.33 (0.47)	0.25 (0.39)	0.29 (0.26)
6- to 7-year-olds	0.42 (0.47)	0.38 (0.52)	0.40 (0.49)	0.67 (0.41)	0.81 (0.39)	0.73 (0.36)	0.29 (0.45)	0.33 (0.47)	0.31 (0.44)
M	0.40 (0.48)	0.44 (0.51)	0.42 (0.47)	0.78 (0.36)	0.78 (0.41)	0.78 (0.33)	0.31 (0.45)	0.29 (0.42)	0.30 (0.35)

(Table 3) reveals that whereas younger children were less accurate with respect to their own actions (hence less willing to acknowledge that they had played with the house), older children were less accurate with respect to the instigator's actions.

Summary. Coaching reliably affected children's tendency to disclose both true and false play when asked repeated yes–no questions. The oath consistently showed positive effects in terms of reducing false denials and false alarms. Reassurance had no effect. Of interest, question repetition did not affect children's accuracy and led very few children to change their responses. Age had virtually no effect on children's performance.

Suppositional Questions

The suppositional questions presupposed that the child and instigator had engaged in play and merely asked the child to clarify details of that interaction. The questions were highly leading and followed the recipient's insistence that the child and the instigator had played with the toy house. This component of the procedure was analogous to suggestibility studies that have elicited high rates of false allegations among young children (e.g., Thompson et al., 1997). We coded children's answers for whether they acquiesced to the presupposition that they had played with the house; acquiescence was scored as correct for children who had played with the house and incorrect for children who had not played with the house. Three percent of children's responses across the questions were either "do not know" or unscorable and were not considered further. Children's accuracy in response to the suppositional questions is presented in Table 4.

An ANOVA on children's accuracy scores with age, instigator, and induction as between-subjects factors and actor (child vs. instigator) as a within-subjects factor yielded significant effects due to age, $F(1, 174) = 7.86$, $p < .01$, partial $\eta^2 = .04$, and instigator, $F(1, 174) = 104.43$, $p < .001$, partial $\eta^2 = .64$. Younger children ($M = 0.53$) were less accurate than older children ($M = 0.64$); children who did not play with the toy house (and for whom the suppositional questions were misleading) were less accurate than children who did, and children who were coached were less accurate than children who were not coached: play $M = 0.92$, warn $M = 0.83$, no-play $M = 0.43$, no-play coach $M = 0.11$; all $ts > 4.21$, $ps < .001$, $ds > .56$.

However, a significant interaction between age and instigator, $F(3, 174) = 6.39$, partial $\eta^2 = .10$, showed that the age effect was limited to the no-play condition, $F(1, 47) = 13.27$, $p < .01$, partial $\eta^2 = .22$. When children neither played with the house nor were coached to claim they had, younger children were significantly less able to resist the questions presupposing house play ($M = 0.24$) than were the older children ($M = 0.62$). Age differences were nonsignificant for children in the other three instigator conditions.

An interaction between actor and instigator, $F(3, 174) = 9.10$, partial $\eta^2 = .14$, $ps < .01$, showed that in the absence of coaching, children were more willing to acquiesce to suggestions about the

Table 4

Means (Standard Deviations) for Proportion Accuracy Scores to Suppositional Questions

	Control ($n = 65$)			Oath ($n = 67$)			Reassurance ($n = 66$)		
	Child	Instigator	*M (SD)*	Child	Instigator	*M (SD)*	Child	Instigator	*M (SD)*
Play ($n = 52$)									
4- to 5-year-olds	0.96 (0.12)	1.00 (0.00)	0.98 (0.06)	0.96 (0.12)	1.00 (0.00)	0.98 (0.06)	0.90 (0.20)	1.00 (0.00)	0.95 (0.10)
6- to 7-year-olds	0.90 (0.20)	1.00 (0.00)	0.95 (0.10)	0.93 (0.14)	0.97 (0.11)	0.95 (0.08)	0.93 (0.21)	1.00 (0.00)	0.97 (0.11)
M	0.93 (0.16)	1.00 (0.00)	0.96 (0.08)	0.94 (0.13)	0.98 (0.08)	0.96 (0.07)	0.92 (0.20)	1.00 (0.00)	0.96 (0.10)
Play coach ($n = 49$)									
4- to 5-year-olds	0.75 (0.46)	0.75 (0.39)	0.75 (0.42)	0.89 (0.33)	0.78 (0.37)	0.83 (0.33)	0.96 (0.12)	0.96 (0.12)	0.96 (0.12)
6- to 7-year-olds	0.81 (0.27)	0.79 (0.40)	0.80 (0.32)	0.79 (0.29)	0.92 (0.24)	0.85 (0.16)	0.71 (0.33)	0.83 (0.25)	0.77 (0.22)
M	0.78 (0.37)	0.77 (0.38)	0.96 (0.08)	0.84 (0.31)	0.84 (0.31)	0.96 (0.07)	0.83 (0.27)	0.90 (0.20)	0.96 (0.10)
No-play ($n = 49$)									
4- to 5-year-olds	0.29 (0.38)	0.19 (0.37)	0.24 (0.37)	0.50 (0.36)	0.19 (0.37)	0.34 (0.35)	0.25 (0.15)	0.00 (0.00)	0.13 (0.08)
6- to 7-year-olds	0.59 (0.43)	0.33 (0.50)	0.46 (0.43)	0.79 (0.35)	0.69 (0.46)	0.74 (0.39)	0.60 (0.45)	0.69 (0.46)	0.65 (0.41)
M	0.45 (0.42)	0.26 (0.44)	0.36 (0.41)	0.65 (0.37)	0.44 (0.48)	0.54 (0.41)	0.43 (0.38)	0.34 (0.47)	0.39 (0.39)
No-play coach ($n = 48$)									
4- to 5-year-olds	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.17 (0.36)	0.17 (0.36)	0.17 (0.36)	0.04 (0.12)	0.00 (0.00)	0.02 (0.06)
6- to 7-year-olds	0.29 (0.45)	0.21 (0.35)	0.25 (0.39)	0.25 (0.39)	0.21 (0.40)	0.23 (0.39)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
M	0.15 (0.34)	0.10 (0.26)	0.13 (0.30)	0.21 (0.36)	0.19 (0.36)	0.20 (0.36)	0.02 (0.08)	0.00 (0.00)	0.01 (0.04)

confederate's play than about their own play. Differences were significant in the play and no-play conditions, $Fs(1, 51 \text{ and } 1, 48) > 7.22, ps \leq .01$, partial $\eta^2 = .12$ and $.23$, respectively, but not in the two coaching conditions. In the play condition, this meant that their accuracy was higher for the instigator's actions ($M = 0.99$) than for their own ($M = 0.91$) because acquiescence reflected accurate responding. Conversely, in the no-play condition, children's greater willingness to acquiesce to questions about the instigator meant that their accuracy was lower for the instigator's actions ($M = 0.35$) than for their own ($M = 0.51$) because acquiescence reduced accuracy.

Summary. The recipient's insistence that the child and confederate had played with the toy, coupled with highly leading questions presupposing that play had occurred, effectively led to very high rates of acquiescence. And, as expected, in the face of such strong suggestion, truth induction had no effects on children's accuracy. Children exhibited greater acquiescence to questions about the confederate's actions than about their own but only in the absence of coaching. We also observed age differences in accuracy but only in the absence of either coaching or play with the toy.

Relations Between Oath-Taking Competency and Performance

The final set of analyses examined whether children's accuracy was related to their performance on the oath-taking competency task, which included questions about children's understanding of both the meaning and the consequences of lying. The task was designed to assess children's basic understanding of the fact that truth and lie refer to factual and counterfactual statements, respectively, and that telling lies has negative consequences, which constitutes the minimum legal standard for qualification to take the oath (Lyon, 2000). Children passing such tasks might not fully appreciate distinctions among jokes, lies, and pretense, and might not have internalized standards against lying but would nevertheless believe that the oath required their best efforts to avoid falsehoods and that falsehoods would be subject to punishment. Children's performance on the two subtasks was significantly related, $r(198) = .42$, and because children would be expected to do well on both tasks in order to qualify to take the oath, we combined children's scores.

Correlations were computed between children's oath-taking competency scores and their free-recall, repeated yes–no question, and suppositional question performance. Because we predicted that the scores would predict performance specifically when children were given the oath (cf. Talwar et al., 2004), we analyzed the truth induction conditions separately. None of the correlations between children's competency scores and interview performance were significant among children in the reassurance and control conditions, $rs = -.05$ to $.11$, and $dfs = 59$ to 63. Among children in the oath condition, however, higher scores on the competency measure were associated with greater accuracy when answering the suppositional questions, $r(67) = .27, p < .05$ (both with respect to actions of the child, $r = .26$, and instigator, $r = .27$, $ps < .05$). Because children's competency performance was correlated with age, $r(198) = .40$, we recomputed the significant relations controlling for age; the correlation between children's competency scores and both their overall performance on the suppositional questions and suppositional questions regarding their own behavior remained significant, $rs(64) = .25$ and $.26$, $ps < .05$, respectively.

A legal perspective suggests an additional approach to analyzing the significance of children's competency scores. In court, competency is a dichotomous judgment such that children who fail to perform well on questions about the meaning and consequences of lying are found incompetent to take the oath and are not allowed to testify. Hence, it is worthwhile to consider if children who failed to perform at ceiling on the competency questions ($N = 61$) were less likely to be influenced by the oath than children who performed at ceiling ($N = 71$; examining children in the control and oath conditions). Hence, we entered children's performance on the oath-taking competency questions (at ceiling vs. not at ceiling) as a between-subjects variable in a series of ANOVAs examining postinduction recall scores, repeated yes–no question accuracy, and suppositional question accuracy. Competency did not emerge as a significant predictor of performance in any of the analyses, nor did competency interact with induction to affect children's performance. Instead, the same truth induction effects already reported reemerged, both among children who had and had not performed at ceiling on the competency measure. In other words, children who failed to pass the competency task were no less likely to be influenced by the oath.

Discussion

The purpose of this study was to examine several potentially critical factors that may influence dishonesty in maltreated children. Our particular focus concerned the role of two types of adult influence: an instigator who extensively coaches a false report and a recipient who asks increasingly suggestive questions with the aid of truth induction. We created a strong coaching manipulation in order to provide a rigorous test of the potential for truth induction to undo coached false reports. We assessed an age range that has been found to undergo substantial development in the proclivity to behave dishonestly and susceptibility to adult influence and tested children's attitudes about the meaning and consequences of lying as a possible correlate of dishonesty.

Instigator Behavior

The coaching was extensive. The instigator provided a rationale for the false story, practiced answering questions with the child, and for the children coached to deny play with the house, practiced a cover story. The recipient questioned the child immediately postcoaching, maximizing the likelihood that the child would remember the coaching and maintain the motivation to respond dishonestly. As predicted, coaching had consistent and robust effects on children's honesty, both when coached to claim that play had occurred and when coached to deny that play had occurred. That is, across analyses, the effect sizes routinely fell in the moderate to large range (Cohen, 1988).

Recipient Behavior

Despite the extensiveness and effectiveness of the coaching, the oath frequently led to improvements in children's performance. Whereas prior research has found that the oath increases children's honesty when they have been warned not to disclose transgressions (Lyon & Dorado, in press; Talwar et al., 2002, 2004), ours is the first to find that the oath increases accuracy among children extensively coached to either conceal or falsely reveal information. For children coached to deny play, the oath exhibited the most consistent positive effects, most of which were moderate in size, increasing disclosure among younger children in their free recall, and among both age groups in response to the repeated yes–no questions. For children coached to claim play falsely, the oath did not affect children's free-recall performance but led to greater accuracy in response to the repeated yes–no questions. Moreover, whereas prior research has found that specific reassurance in which the recipient explicitly mentions the target behavior can impair performance (Lyon & Dorado, in press), this study demonstrates that general reassurance exhibited no reliable tendency to increase false reports. At the same time, however there was only very limited evidence that reassurance improved the accuracy of children's reports.

Repeated yes–no questioning, often characterized as highly suggestive (Ceci & Friedman, 2000; Endres, Poggenpohl, & Erben, 1999), did not undermine truth induction's efficacy. Indeed, we were surprised to find that repeated questioning had no effect on children's accuracy, and very few children changed their responses. This was so despite the fact that the questions were repeated immediately and with a skeptical tone, thus maximizing their suggestiveness (Endres et al., 1999). It may be that children's certainty about what actually occurred undermined repetition's effect (Lyon, 2002).

Truth induction's benefits did diminish, however, in response to the highly leading suppositional questions. When the recipient insisted that play with the house had occurred and asked a series of questions that presupposed play, false reports were high, a common finding in the suggestibility literature (e.g., Bruck, Ceci, & Principe, 2006). Under these circumstances, neither the oath nor reassurance had any effect. The results thus suggest that administering the oath in conjunction with free-recall and yes – no questions improves young children's performance but that any benefits of truth induction may be overridden if highly leading questions are asked. Moreover, although the differences were not statistically significant, reassured children occasionally looked worse than children in the control group. This pattern is consistent with other research finding that normally benign interviewing techniques may be contraindicated if coupled with highly suggestive questioning (e.g., Bruck, Melnyk, & Ceci, 2000).

Child Characteristics

We found some support for our prediction that older children would be more affected by coaching, with respect to both concealing true play and claiming false play. However, these age differences did not appear in response to yes – no and suppositional questions. With respect to truth induction, on the other hand, there was virtually no evidence that age mattered; indeed, the one age difference was contrary to our prediction, in that younger children were more influenced than older children by truth induction in free recall. We had anticipated that younger children's limited appreciation of the consequences of disclosure and the morality of lying would lead them to be less responsive to truth induction, but there was little evidence that this was the case. Rather, the most consistent finding was that children in this age range were quite uniform in their susceptibility to recipient influence.

Finally, there were few significant (and no large) age differences in overall accuracy, even when one would expect age differences due to suggestibility among younger children. Younger children showed some tendency to initially provide less information overall, which is consistent with the difficulty they have in spontaneously producing information when asked for free recall. However, there were no age differences in accuracy in response to the repeated yes – no questions. Younger children showed a greater tendency to acquiesce to the highly suggestive suppositional questions than older children, a common finding in suggestibility research (Bruck et al., 2006), but this was true in only one of the four instigator conditions. The lack of age effects in accuracy might be attributable at least in part to the fact that because we were interested in honesty and dishonesty, we deliberately minimized memory errors by questioning children immediately after their interaction with the instigator. Had a longer delay taken place between the instigator's instructions and the recipient's questioning, age differences may have emerged.

We also predicted that children's understanding of the meaning and consequences of lying would be positively related to their honesty in the oath condition. There were indeed some significant correlations, although they were infrequent. Moreover, the additional analyses suggested by legal approaches to competency suggest that although the competency scores have some predictive value in assessing the accuracy of children sworn to tell the truth, a dichotomous yes – no judgment about competency based on children's performance is likely to exclude children from testifying who would nevertheless be influenced by the oath (cf. Talwar et al., 2002, who found evidence that a promise to tell the truth decreased dishonesty among children who would probably fail an oath-taking competency assessment).

Age differences might nevertheless emerge if the age range is increased. Of course, still younger children will at some point simply lack the necessary comprehension to be influenced by coaching or a promise to tell the truth. Older children might be less influenced by either coaching or truth induction, because their moral behavior is increasingly influenced by self-evaluation as opposed to external influences (Bandura, 1991). Moral self-evaluations undergo substantial development during the early grade school years. Bussey (1992, 1999) found that it was not until second grade that children anticipated more pride for telling the truth than for lying, and some of the research has found that older grade schoolchildren are less inclined than younger children to lie for a transgressing instigator (Bussey & Grimbeek, 1995; Pipe & Wilson, 1994; but see Talwar et al., 2004).

It will be necessary, in subsequent research, to better identify individual differences in children's response to truth induction. Those individual differences, in turn, will give us insight into the mechanism by which truth induction exerts its effect. For example, if older children are less influenced by truth induction and more influenced by internal moral standards, this suggests that truth induction affects children's perceptions of the costs and benefits of honesty. Future research should directly assess children's expectations about the consequences of nondisclosure and disclosure, dishonesty and honesty.

Limitations

Although the present study's results are provocative and highlight the potential value of truth induction in decreasing dishonesty among maltreated children, several limitations should be acknowledged. First, the experimental context necessarily limited the external validity of the study. On the one hand, child witnesses obviously face much more serious pressures about more serious transgressions. Maltreated children often fear quite severe consequences for disclosure (Anderson et al., 1993), which may be little affected by reassurances regarding the recipient's reactions. A promise to tell the truth might be taken more seriously in the real world, but it is still likely insufficient to overcome many pressures. On the other hand, it may be easier to motivate honesty in children questioned about more serious events, given children's desire to avoid harm to themselves or others. Both false allegations and false denials carry more serious consequences in the real world than in our research.

Second, a similar and equally serious problem with external validity is that both the instigator and the recipient in this study were virtual strangers to the participants. Of course, because our participants were maltreated children who had been removed from their parents' custody, we could not examine the power of parents as instigators or recipients in influencing children's dishonesty. Recent research on children's disclosure of sexual abuse, however, demonstrates quite clearly the potential influence of adults who are emotionally close to the child (Hershkowitz, Lanes, & Lamb, 2007; Malloy et al., 2007).

Third, children in the coaching conditions were not truly "lying" but "tricking" the recipient, a necessary modification given the ethical problems inherent in teaching children to lie. A trick is intended to deceive but is only a temporary lie, as manifested by the instigator's reassurance to the child that they would ultimately disclose what they had done. At the same time, the high rates of nondisclosure and denial of play among the children who received no coaching suggested that they intuited that play with the toy house constituted a transgression (however minor), and these children had no expectation that the truth would be revealed.

Future Directions

It may be possible to increase the efficacy of truth induction in reducing dishonesty. Reassurance might be made effective if the recipient reassures the child about the reactions of parties other than the recipient, although in the real world, such reassurances might be difficult to make with confidence. A combination of reassurance with a promise to tell the truth might also have benefits. The combination could be particularly effective if children who have been coached to lie might otherwise interpret reassurance as guaranteeing that a false story will not be punished. When coached false reports are a concern, truth induction might be coupled with inquiries about coaching itself. Those questions might be quite pointed ("Did the man tell you what to say?") or more general ("What did the man say to you?"). The recipient in this study only pursued the hypothesis that the child had engaged in play with the toy house, making it incumbent on the child to disclose coaching.

Further work can also elucidate the conditions under which instigators exert stronger or weaker effects. In this study, the instigator was himself a transgressor who engaged the child in the transgression (recall that the coached children were either told that they were supposed to play, in the no-play coach condition, or that they were not supposed to play, in the play coach condition). We deliberately chose this design to mimic situations in which children themselves feel implicated in adults' wrongdoing (a common dilemma for many sexually abused children; Anderson et al., 1993). However, whether the instigator is the same person as the potential wrongdoer and whether the child is herself implicated in the target behavior are important factors to manipulate in subsequent research.

It will also be important in future to directly compare maltreated and nonmaltreated children. Instigator and recipient influences may have differential effects depending on a child's maltreatment history. Maltreatment is likely to have negative effects on children's trust that adults will not cause them harm (Shields, Ryan, & Cicchetti, 2001), which could make them particularly susceptible to secrecy. At the same time, developmental delays might make them less proficient at lying. Direct comparisons will enable us to better understand the effects of maltreatment on children's attitudes and behavior, the mechanisms affecting children's dishonesty, and the implications of child witness research for actual practice.

Finally, truth induction should be attempted in the field. Our focus on maltreated children who both had and had not transgressed highlights the study's implications for forensic interviewers assessing the costs and benefits of various interviewing approaches. Children who testify are disproportionately those who have been substantiated as maltreated and whose honesty is subject to challenge. Before truth induction can safely be attempted in the field, where its ultimate utility will be determined, it was necessary to establish its effects on both true and false reports in a laboratory context, where the truth is known. It was also imperative to examine truth induction in precisely the population that will be the subject of its use in the field. This study lays important groundwork for continued research concerning truth induction in actual forensic interviews.

References

Anderson, J., Martin, J., Mullen, P., Romans, S., & Herbison, P. (1993). Prevalence of childhood sexual abuse experiences in a community sample of women. *Journal of the American Academy of Child and Adolescent Psychiatry, 32,* 911–919.

Bandura, A. (1991). Social cognitive theory of moral thought and action. In W. M. Kurtines & J. L. Gewirtz (Eds.), *Handbook of moral behavior and development* (pp. 45–103). Hillsdale, NJ: Erlbaum.

Bottoms, B. L., Goodman, G. S., Schwartz-Kenney, B. M., & Thomas, S. N. (2002). Understanding children's use of secrecy in the context of eyewitness reports. *Law & Human Behavior, 26,* 285–313.

Brennan, M. (1994). The battle for credibility: Themes in the cross-examination of child victim witnesses. *International Journal for the Semiotics of Law, 26,* 51–73.

Bruck, M., Ceci, S. J., & Principe, G. F. (2006). The child and the law. In K. A. Renninger, I. E. Sigel, W. Damon, & R. M. Lerner (Eds.), *Handbook of child psychology* (6th ed., pp. 776–816). Hoboken, NJ: John Wiley & Sons.

Bruck, M., Melnyk, L., & Ceci, S. J. (2000). Draw it again Sam: The effect of drawing on children's suggestibility and source monitoring ability. *Journal of Experimental Child Psychology, 77,* 169–196.

Bussey, K. (1992). Lying and truthfulness: Children's definitions, standards, and evaluative reactions. *Child Development, 63,* 129–137.

Bussey, K. (1999). Children's categorization and evaluation of different types of lies and truths. *Child Development, 70,* 1338–1347.

Bussey, K., & Grimbeek, E. J. (1995). Disclosure processes: Issues for child sexual abuse victims. In K. J. Rotenberg (Ed.), *Disclosure processes in children and adolescents* (pp. 166–203). Cambridge, England: Cambridge University Press.

Bussey, K., Lee, K., & Grimbeek, E. J. (1993). Lies and secrets: Implications for children's reporting of sexual abuse. In G. S. Goodman & B. L. Bottoms (Eds.), *Child victims, child witnesses: Understanding and improving testimony* (pp. 147–168). New York: Guilford.

Ceci, S. J., & Friedman, R. D. (2000). The suggestibility of children: Scientific research and legal implications. *Cornell Law Review, 86,* 34–108.

Ceci, S. J., & Leichtman, M. (1992). "I know that you know that I know that you broke the toy": A brief report of recursive awareness among 3-year-olds. In S. J. Ceci, M. Leichtman, & M. Putnick (Eds.), *Cognitive and social factors in early deception* (pp. 1–9). Hillsdale, NJ: Lawrence Erlbaum.

Cohen, J. (1988). *Statistical power analyses for the social sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

DePaulo, B. M., Jordan, A., Irvine, A., & Laser, P. S. (1982). Age changes in the detection of deception. *Child Development, 53,* 701–709.

Ekman, P. (1989). Why kids lie: How parents can encourage truthfulness. New York: Scribner.

Endres, J., Poggenpohl, C., & Erben, C. (1999). Repetitions, warnings and video: Cognitive and motivational components in preschool children's suggestibility. *Legal and Criminological Psychology, 4,* 129–146.

Feben, D. J. (1985). *Age of witness competency: Cognitive correlates.* Unpublished honours thesis, Monash University, Clayton, Australia.

Goodman, G. S., Aman, C. J., & Hirschman, J. (1987). Child sexual and physical abuse: Children's testimony. In S. J. Ceci, M. P. Toglia, & D. F. Ross (Eds.), *Children's eyewitness memory* (pp. 1–23). New York: Springer-Verlag.

Goodman, G. S., Batterman-Faunce, J. M., Schaaf, J. M., & Kenney, R. (2002). Nearly 4 years after an event: 'Children's eyewitness memory and adults' perceptions of children's accuracy. *Child Abuse & Neglect, 26,* 849–884.

Hershkowitz, I., Lanes, O., & Lamb, M. E. (2007). Exploring the disclosure of child sexual abuse with alleged victims and their parents. *Child Abuse & Neglect, 31,* 111–123.

Huffman, M. L., Warren, A. R., & Larson, S. M. (1999). Discussing truth and lies in interviews with children: Whether, why, and how? *Applied Developmental Science, 1,* 6–15.

Lewis, M., Stanger, C., & Sullivan, M. W. (1989). Deception in 3-year-olds. *Developmental Psychology, 25,* 439–443.

London, K., & Nunez, N. (2002). Examining the efficacy of truth/lie discussions in predicting and increasing the veracity of children's reports. *Journal of Experimental Child Psychology, 83,* 131–147.

Lyon, T. D. (2000). Child witnesses and the oath: Empirical evidence. *Southern California Law Review, 73,* 1017–1074.

Lyon, T. D. (2002). Applying suggestibility research to the real world: The case of repeated questions. *Law and Contemporary Problems, 65,* 97–126.

Lyon, T. D., & Dorado, J. S. (in press). Truth induction in young maltreated children: The effects of oath-taking and reassurance on true and false disclosures. *Child Abuse & Neglect.*

Lyon, T. D., & Saywitz, K. J. (1999). Young maltreated children's competence to take the oath. *Applied Developmental Science, 3,* 16–27.

Malloy, L. C., Lyon, T. D., & Quas, J. A. (2007). Filial dependency and recantation of child sexual abuse allegations. *Journal of the American Academy of Child & Adolescent Psychiatry, 46,* 162–170.

Peterson, C. C., Peterson, J. L., & Seeto, D. (1983). Developmental changes in ideas about lying. *Child Development, 54,* 1529–1535.

Pipe, M., & Wilson, J. C. (1994). Cues and secrets: Influences on children's event reports. *Developmental Psychology, 30,* 515–525.

Polak, A., & Harris, P. L. (1999). Deception by young children following noncompliance. *Developmental Psychology, 35,* 561–568.

Quas, J. A., Davis, E., Goodman, G. S., & Myers, J. E. B. (2007). Repeated questions, deception, and children's true and false reports of body touch. *Child Maltreatment, 12,* 60–67.

Sas, L. D., & Cunningham, A. H. (1995). *Tipping the balance to tell the secret: The public discovery of child sexual abuse.* London, Ontario: London Family Court Clinic.

Shields, A., Ryan, R. M., & Cicchetti, D. (2001). Narrative representations of caregivers and emotional dysregulation as predictors of maltreated children's rejection by peers. *Developmental Psychology, 37,* 321–337.

Shonk, S. M., & Cicchetti, D. (2001). Maltreatment, competency deficits, and risk for academic and behavioral maladjustment. *Developmental Psychology, 37,* 3–17.

Stouthamer-Loeber, M., & Loeber, R. (1986). Boys who lie. *Journal of Abnormal Child Psychology, 14,* 551–564.

Summit, R. C. (1983). The child sexual abuse accommodation syndrome. *Child Abuse & Neglect, 7,* 177–193.

Talwar, V., Gordon, H., & Lee, K. (2007). Lying in the elementary school years: Verbal deception and its relation to second-order belief understanding. *Developmental Psychology, 43,* 804–810.

Talwar, V., & Lee, K. (2002). Development of lying to conceal a transgression: Children's control of expressive behaviour during verbal deception. *International Journal of Behavioral Development, 26,* 436–444.

Talwar, V., Lee, K., Bala, N., & Lindsay, R. C. L. (2002). Children's conceptual knowledge of lying and its relation to their actual behaviors: Implications for court competence examinations. *Law & Human Behavior, 26,* 395–415.

Talwar, V., Lee, K., Bala, N., & Lindsay, R. C. L. (2004). Children's lie-telling to conceal a parent's transgression: Legal implications. *Law & Human Behavior, 28,* 411–435.

Tate, C. S., Warren, A. R., & Hess, T. M. (1992). Adults' liability for children's "lie-ability": Can adults coach children to lie successfully? In S. J. Ceci, M. D. Leichtman, & M. Putnick (Eds.), *Cognitive and social factors in early deception* (pp. 69–87). Hillsdale, NJ: Erlbaum.

Thompson, W. C., Clarke-Stewart, A., & Lepore, S. J. (1997). What did the janitor do? Suggestive interviewing and the accuracy of children's accounts. *Law & Human Behavior, 21,* 405–426.

Tye, M. C., Amato, S. L., Honts, C. R., Devitt, M. K., & Peters, D. (1999). The willingness of children to lie and the assessment of credibility in an ecologically relevant laboratory setting. *Applied Developmental Science, 3,* 92–109.

Wilson, J. C., & Pipe, M. (1989). The effects of cues on young children's recall of real events. *New Zealand Journal of Psychology, 18,* 65–70.

Wilson, A. E., Smith, M. D., & Ross, H. S. (2003). The nature and effects of young children's lies. *Social Development, 12,* 21–45.

Child Development, July/August 2008, Volume 79, Number 4, Pages 930 – 943

Genetic Influences on Activity Level in Early Childhood: Do Situations Matter?

Kimberly J. Saudino
Boston University

Jeffrey A. Zapfe
Acentech Inc.

Although genetic influences on individual differences in activity level (AL) are well documented, few studies have considered the etiology of AL from a contextual perspective. In the present study, cross-situational and context-specific genetic effects on individual differences in AL at age 2 were examined. The AL of 312 twin pairs (144 monozygotic and 168 dzygotic) was mechanically assessed with actigraphs in the home and in laboratory test and play situations. AL displayed significant genetic variance in all 3 situations. Moreover, actigraph scores significantly correlated across situations. Multivariate genetic model-fitting analyses found that the observed cross-situational continuity in AL was due entirely to genetic factors. Situational differences in AL arise from genetic, shared environmental, and nonshared environmental influences.

Activity level (AL) is a pivotal individual differences variable in infancy and early childhood. A core dimension of nearly every temperament theory, AL is a highly salient feature of child behavior that has been linked to various facets of child health and development. Twin studies of temperament in infancy and early childhood have yielded considerable evidence that AL is genetically influenced (e.g., Braungart, Plomin, DeFries & Fulker, 1992; Cyphers, Phillips, Fulker, & Mrazek, 1990; Saudino & Eaton, 1991, 1995; Saudino, Plomin, & DeFries, 1996; Stevenson & Fielding, 1985); however, this is only a first step in the understanding of the etiology of this important temperament dimension. There are many questions that remain unanswered. For example, although developmentalists frequently acknowledge the importance of contextual factors on temperament (e.g., Mangelsdorf, Schoppe, & Burr, 2000), short-term situational changes in AL have rarely been considered from a behavioral genetic perspective. Because temperament is viewed as being stable, questions of genetic and environmental contributions to situational change have largely been ignored by temperament researchers. However, different situations place different demands on the individual and elicit different behaviors.

Longitudinal studies of developmental changes in AL suggest that genetic factors contribute to stability, whereas change is due to environmental factors (Saudino & Cherny, 2001a; Saudino et al., 1996), but the factors that influence developmental change and short-term situational change need not be the same (Plomin & Nesselroade, 1990). Behavioral change across situations is generally assumed to be due to environmental differences; however, this is an empirical question. It is possible that genetic factors contribute to short-term change as the individual goes from situation to situation. That is, individual differences in responding to specific situations might be influenced by genetic effects (Phillips & Matheny, 1997).

There is little doubt that AL is influenced by situational factors. The types of toys available, the presence of peers, and the stressfulness, novelty, or restrictiveness of the situation can affect children's mean levels of AL (e.g., Eaton & Keats, 1982; Fagot & O'Brien, 1994; Pinto & Tryon, 1996). More specific to the question of individual differences and cross-situational consistency and change, there is also change in terms of one's ordinal position within the group. For example, observations of AL at home correlate only modestly with observations of AL in playgroup situations or at school (Fagot & O'Brien, 1994; Schaughency & Fagot, 1993). Similarly, correlations between mechanical measures of AL taken in free play and test settings tend to be low in both clinical and control samples (Barkley & Ullman, 1975). Thus, the child who is active in one situation is not necessarily active in another. If this is the case, there must be some difference in the genetic and/or

---

The Boston University Twin Project (BUTP) is supported by grant MH062375 from the National Institute of Mental Health. The authors gratefully acknowledge the parents and twins in the BUTP. We also thank Dr. Michael Neale for his generous assistance with Mx.

Correspondence concerning this article should be addressed to Kimberly J. Saudino, Psychology Department, Boston University, 64 Cummington Street, Boston, MA 02215. Electronic mail may be sent to ksaudino@bu.edu.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7904-0009

environmental influences on individual differences in AL across situations.

Detecting Genetic Influences on Situational Change

The etiology of cross-situational consistency and change in temperament can be explored by using multivariate genetic analyses that assess the extent to which genetic effects on a measure of temperament in one situation are shared with genetic effects on another measure of the same dimension in a different situation. *Cross-situational* genetic effects are genetic influences that overlap across different situations. Such situation-general genetic factors contribute to the stability of temperament across situations. *Contextual* genetic influences are genetic effects that are unique to a specific situation and, consequently, contribute to situational changes in temperament. For example, in the MacArthur Longitudinal Twin Study, multivariate genetic analyses have been used to examine the extent to which genetic effects on a measure of shyness in the laboratory overlap with shyness in the home environment (Cherny, Fulker, Corley, Plomin & DeFries, 1994; Cherny et al., 2001). Although shyness in the laboratory and home were only modestly correlated, there was substantial overlap between the genetic influences involved in both situations. This finding of cross-situational genetic effects for shyness in the laboratory and the home means that genetic factors contribute to the stability of shyness across the two situations. Indeed, the observed phenotypic correlation between the two measures of shyness was almost entirely due to overlapping genetic effects. In contrast, differences between shyness in the laboratory and in the home were due primarily to environmental factors, although there were some small contextual, or situationally specific, genetic influences on shyness in the home.

What is the role of genetic influences on situational change in AL? Although not well studied, prior research suggests that genetic factors may contribute to *both* cross-situational stability and situational change in AL. For example, multivariate genetic analyses of teacher and tester-rated temperament in the Colorado Adoption Project found that although teacher and tester ratings of AL at age 7 were only modestly correlated (r = .19), there was substantial genetic overlap between the two measures (Schmitz, Saudino, Plomin, Fulker, & DeFries, 1996). Forty-six percent of the genetic variance in tester ratings was shared with teacher ratings. Moreover, only genetic factors contributed significantly to the covariance between teacher and tester ratings, indicating that what teachers and testers see in common in children's AL is due to genetic influences. Despite this evidence of cross-situational genetic effects, there was also significant genetic influence unique to tester ratings, indicating context-specific genetic effects for activity (i.e., 54% of the genetic variance on tester ratings was independent of genetic variance on teacher ratings). Similar results have emerged from an application of a multitrait–multimethod approach to infant twin data comprising three temperament dimensions (activity, emotional tone, and sociability) assessed three different ways (tester rating, playroom observation, and parent rating) that found substantial context-specific genetic variance in addition to cross-situational genetic variance (Phillips & Matheny, 1997).

These two studies are a first step in exploring genetic and environmental contributions to short-term situational change; however, in both studies, situational variance is confounded with method variance. For example, in the Schmitz et al. (1996) study, the informants differed across situation (i.e., the test and school situations). Similarly, in the Phillips and Matheny (1997) study, both informants *and* measures differed across the three situations. Research has clearly shown that for ratings of temperament and/or behavior problems, different informants typically do not contribute the same information about the child's behavior (Achenbach, McConaughy, & Howell, 1987; Mangelsdorf et al., 2000). Therefore, the confounding of situational and method variance in these studies limits conclusions regarding cross-situational consistency and change because it is difficult to say whether the context-specific genetic variance refers to true situational change or method differences.

Disentangling situational variance from method variance. To understand true situational change, we must disentangle situational variance from method variance. To do this, it is necessary to assess the behavior of interest with the *same* measure across different situations. For some dimensions of temperament (e.g., emotionality), this is difficult because different informants are often needed for different situations (e.g., home and school). This, however, is not a problem with AL because AL can be mechanically assessed across a variety of situations.

Despite the advantages of mechanical measures of AL, only one prior study has used mechanical motion recorders to explore situational effects in the etiology of individual differences in AL. A sample of school-aged twins wore motion recorders in a laboratory cognitive testing situation and during a 25-min break in testing (Wood et al., 2007). AL in the two situations was substantially correlated (r = .52), and AL was significantly heritable in both situations (h^2 test = .24, h^2 break = .30). In contrast to prior work suggesting contextual or situationally specific genetic effects (i.e., Phillips & Matheny, 1997; Schmitz et al., 1996), the genetic correlation between the two situations was estimated as 1.0, indicating complete overlap between the genetic influences operating on the two situations. At question, however, is whether the two laboratory situations were sufficiently different to reasonably expect that they would have different genetic etiologies. That is, more diverse situations may produce different results.

In the present study, cross-situational and context-specific genetic effects on individual differences in AL at age 2 were examined using mechanical motion recorders to assess AL in the home as well as in the laboratory. Twins wore motion recorders for 2 days in their homes, thus providing a naturalistic measure of AL. In addition, two laboratory-based measures of AL were obtained: one taken during a standardized test situation and another based on the child's behavior in a series of play-based activity episodes. Although the testing situation is a limited context in which to observe children's behavior, it is a standard situation that facilitates comparisons among children. Moreover, the somewhat stressful nature of the test situation enriches observations of certain temperament-like behaviors. In contrast, the laboratory play situation allows us to evaluate behaviors that may be more typical of children's day-to-day activity. Play situation episodes are standard across children, but activity within each episode is directed by the child. Thus, the present study explores situationally specific genetic effects in the home and in the laboratory situations and between the laboratory test and the play situations. Although we made no specific predictions about differential heritability across situations, it was predicted that AL would be genetically influenced in each of the three situations. Further, based on prior research and temperament theory, it was hypothesized that cross-situational continuity would be largely due to genetic factors (i.e., genetic influences contribute to the phenotypic correlation between situations). Nonetheless, given that laboratory measures of AL are typically modestly correlated with AL in the home, it was also predicted that there would be some context-specific genetic effects.

Method

Sample

The Boston University Twin Project sample was recruited from birth records supplied by the Massachusetts Registry of Vital Records. Twins were selected preferentially for higher birth weight and gestational age. No twins with birth weights less than 1,750 g or with gestational ages less than 34 weeks were included in the study. Twins were also excluded if one or both twins had a health problem that might affect motor activity (e.g., cerebral palsy, club foot) or had chromosomal abnormalities. The present analyses include 312 same-sex pairs of twins (144 monozygotic [MZ] and 168 dzygotic [DZ]), mean age 2.07 years (*SD* = 0.05). Although the sample was predominately Caucasian (85.4%), ethnicity was generally representative of the Massachusetts population (3.2% Black, 2% Asian, 7.3% mixed, and 2.2% other). Socioeconomic status according to the Hollingshead (1975), four-factor index ranged from low to upper middle class (range = 20.5 – 66; *M* = 50.9, *SD* = 14.1).

Zygosity was determined via DNA analyses using DNA obtained from cheek swab samples. In the cases where DNA was not available (*n* = 3), zygosity was determined using parents' responses on physical similarity questionnaires, which have been shown to be more than 95% accurate when compared to DNA markers (Price, Freeman, Craig, Ebersole, & Plomin, 2000). In our present sample, we were able to assign zygosity with certainty to 99% of the twin pairs using the parent questionnaire; moreover, agreement between questionnaire and DNA zygosity analyses was very high (κ = .94).

Procedure Overview

The twins were assessed within approximately 2 weeks of their second birthday. The procedure consisted of two visits, 48 hr apart, to the laboratory. At the initial visit, the motion recorders were attached to each child. After attachment of the motion recorders, one twin was assessed within a standardized test situation while the other twin was assessed within the laboratory play situation (both described subsequently). The second visit was scheduled for 48 hr following the attachment of the motion recorders. At this visit, the situations were reversed for each twin. Motion recorders were then removed, and the record sheets and cheek scrapings collected. Twins were assessed by different testers; however, for each twin, the tester was the same across the two laboratory situations. In the present study, the term *home situation* refers to all time spent outside the laboratory.

Test situation. The administration of the mental scale of *Bayley Scales of Infant Development – Second Edition* (Bayley, 1993) was used to provide a structured test situation for assessing AL. Assessment of AL during administration of the Bayley mental scale has been widely used in behavioral genetic research, but prior research has relied solely on observer ratings within this context (e.g., Braungart et al. 1992; Matheny, 1980, 1983; Saudino et al., 1996).

Play situation. Activity episodes from the *Laboratory Temperament Assessment Battery – Preschool Version* (Lab – TAB; Goldsmith, Reilly, Lemery, Longley, & Prescott, 1995) were used to provide standard play situations for assessing AL. Four Lab – TAB activity episodes were included in the present study: Corral of Balls: the child plays in a corral enclosure filled with many brightly colored balls (3 min); Arc of Toys: the child is presented with a variety of age-appropriate toys and is permitted to play with whatever toys he or she chooses (5 min); Workbench: in this measure of fine motor activity, the child is shown a work bench toy and is permitted to play with it (3 min); and Fidgeting Video: the child is seated in a child chair in front of a television monitor and shown an age-appropriate video (4 min).

Measures

Mechanical assessment of AL. During the 2-day data collection period, AL was assessed with the Mini Mitter actical (actigraph). This device is a miniature omnidirectional accelerometer that has been designed to detect frequency and intensity of motion within the range of normal human movement (Mini Mitter Company, 2005). Actigraphs have been shown to be valid and reliable instruments for recording the activity of children (Freedson et al., 1997; Janz, 1994). Each twin was randomly assigned four actigraphs, one per limb. Actigraphs were attached by means of Tyvek adhesive wristbands. Arm attachment, at the wrist, was on the dorsal aspect of the forearm proximal to the radial carpal joint. Leg attachment, at the ankles, was superior to the lateral malleoli.

Parents were provided with instructions regarding the care of the motion recorders and a sheet for logging the times when the actigraphs were off a limb. To differentiate actigraphs within twin pair, each actigraph was covered with white surgical tape on which the child's name was printed. Actigraphs were also color coded and keyed to a parent record sheet indicating the color of actigraph that went on each limb. Thus, the parent was able to easily identify the set of actigraphs that belongs to each twin and, further, to match the instrument to the correct limb. Parents were encouraged to engage in normal activities and to maintain daily routines with their children during the 2-day data collection period. The average length of time that the motion recorders were worn was 48.22 hr ($SD = 3.22$).

To adjust for variations in the total time that each instrument was worn within the home and lab situations, the number of activity units was converted to a rate per minute real time. Similar adjustments were made to the data from the test and play situations. Arm and leg activity counts were highly correlated ($r = .69$, $p < .001$); therefore, composite actigraph scores reflecting overall motor activity for each situation were calculated by averaging the four limb actigraph scores. Aggregating across motion recorders has been shown to increase both the reliability and the validity of AL measures (e.g., Eaton, 1983; Melanson & Freedson, 1995). Consistent with prior research, actigraph scores were positively skewed (Saudino & Eaton, 1991, 1995; Wood et al., 2007) and were square root transformed to create a more normal distribution.

Correlational Analyses

Twin intraclass correlations indexing co-twin similarity were calculated using the double-entry method. Because twin correlations can be inflated by variance due to gender, all scores were residualized for gender effects (McGue & Bouchard, 1984). These residualized scores were used in all behavioral genetic analyses. To evaluate genetic and environmental influences to cross-situational continuity, cross-twin cross-situation correlations were calculated. Cross-twin correlations are the essence of a multivariate analysis of covariance. For the present analyses, the cross-twin cross-situation correlation involved correlating the score of Twin 1 for AL in one situation with score of Twin 2 for AL in another and vice versa. Genetic contributions to the covariance between two measures are implied when the MZ cross-twin correlation is greater than the DZ cross-twin correlation.

Model-Fitting Analyses

Although the major results of a multivariate twin analysis can be gleaned from twin cross-correlations, model-fitting procedures analyze all the data simultaneously, provide tests of the fit of models, yield confidence intervals for parameter estimates, and tests the fit of alternative models (Plomin, DeFries, McClearn, & McGuffin, 2001). Models were fit to raw data using a maximum likelihood pedigree approach implemented in Mx structural equation modeling software (Neale, Boker, Xie, & Maes, 2006). This approach does not yield a chi-square statistic for assessing the fit of the model; however, the fit of a model can be assessed by calculating twice the difference between the negative log likelihood (−2LL) of the model and that of a saturated model (i.e., a model in which the variance–covariance structure is not estimated and all variances and covariances for MZ and DZ twins are estimated). The difference in −2LL is asymptotically distributed as a chi-square statistic with degrees of freedom equal to the difference in the number of parameters in the full model and that in the saturated model.

We fit two multivariate models to the data to examine cross-situational and contextual genetic effects on mechanically assessed AL. Both models decompose the phenotypic variance of each situation and the phenotypic covariances *between* situations into additive genetic effects (A), shared environmental effects (C), and nonshared environmental effects (E); however, each model makes different assumptions about the reasons for consistency across situations. In both models, additive genetic influences refer to the sum of the average effect of all genes that influence a trait. Based on the degree of genetic relatedness, the A factors correlate 1.0 and .5 for MZ and DZ twins, respectively. The C factors refer to the influence of shared rearing environments on twin resemblance. Because all twins were reared in the same family, shared environments correlate 1.0 for both MZ and DZ twins. Finally, the E factors reflect nonshared environmental influences that are unique to each member of a twin pair, including measurement error.

Independent pathway model. The independent pathway model (Kendler, Heath, Martin, & Eaves, 1987) assumes that genetic and environmental factors contribute to covariance between situations through separate genetic and environmental pathways. This model decomposes the genetic, shared environmental, and nonshared environmental variance of AL in the home, test situation, and play situations into genetic and environmental effects that are common to all three measures (i.e., cross-situational effects) and genetic and environmental effects that are unique to each situation (i.e., contextual effects). For example, in Figure 1, the latent variables A, C, and E represent additive genetic, shared environmental, and nonshared environmental effects, respectively, that are common to all three situations. The latent variables denoted in subscripted lowercase represent genetic and environmental effects on AL that are unique to a specific situation (e.g., a_h, c_h, and e_h refer to genetic and environmental influences unique to the home situation). Under this model, covariance between situations can arise due to different factors. That is, although measures of AL may be intercorrelated across all situations, the sources of covariance between any two situations may differ (e.g., the correlation between home and test situations could have a different source than the correlation between the test and the play situations). This model is considered to be "agnostic" in that it does not specify that the same phenotype operates across different situations; rather, it just allows that the phenotypes being assessed in each situation are correlated (Hewitt, Silberg, Neale, Eaves, & Erikson, 1992).

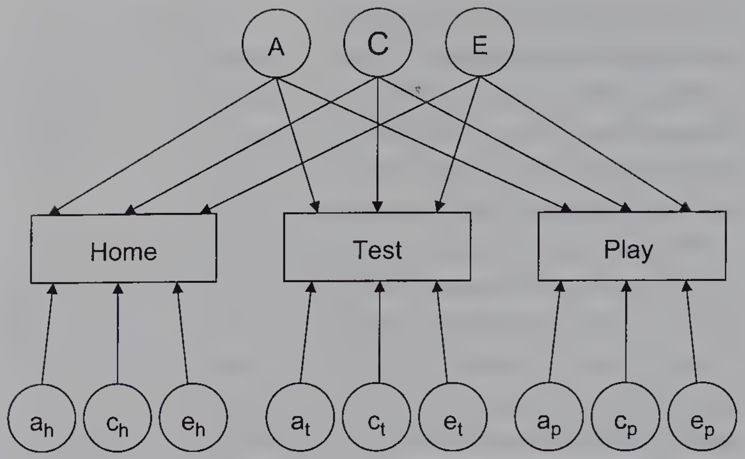


Figure 1. Independent pathway model.

Common pathway model. As compared to the independent pathway model, the common pathway model (Kendler et al., 1987; McArdle, & Goldsmith, 1990) is more restrictive positing that genetic and environmental factors influence covariation between situations through a *single* common pathway. This model suggests that correlations between situations arise because a common phenotype (AL in Figure 2) operates across all situations. This common phenotype is then influenced by genetic and environmental influences. As is the case for the independent pathway model, this model also allows genetic and environmental effects specific to each situation. Under this model, genetic and environmental sources of covariance are the same across all situations. Cross-situational consistency arises because different situations tap the same latent phenotype (i.e., they are assessing the same behaviors). Differences between situations arise because, to some extent, different contexts also evoke different phenotypes.

Results

Descriptive Statistics

Table 1 lists the means and standard deviations by sex and zygosity. We evaluated the mean differences

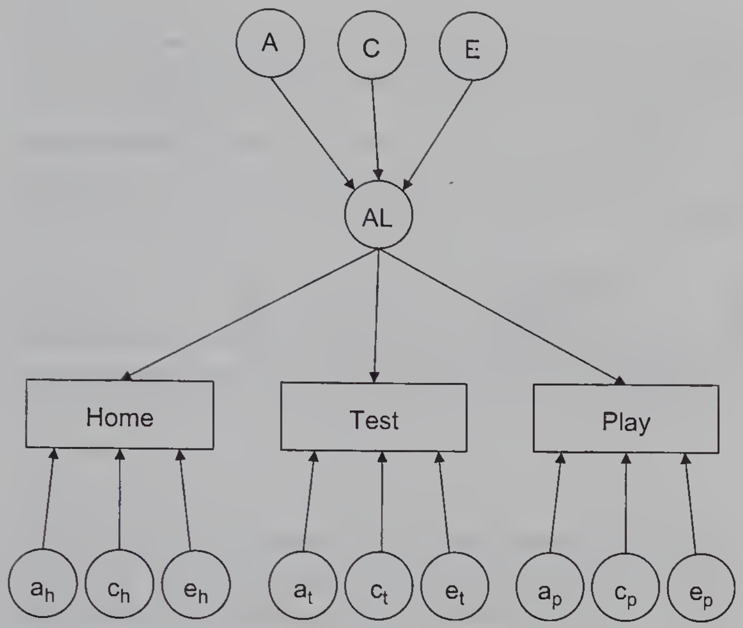


Figure 2. Common pathway model.

using generalized estimating equations (GEE) implemented in the SAS GENMOD procedure to account for dependence in the data due to the fact that our sample comprised pairs of twins. GEE are an extension of the standard generalized linear models that allow modeling of correlated data (Liang & Zeger, 1986; Zeger & Liang, 1986). Consistent with previous research on mechanically assessed AL (Saudino & Eaton, 1991, 1995), males were significantly more active that females (home: $z = -2.32$, $p < .05$; test: $z = -5.27$, $p < .001$; and play: $z = -4.08$, $p < .001$). The effect sizes as estimated by Cohen's *d* (Cohen, 1988) ranged from .24 to .48, indicating that the means for males and females differed by approximately one quarter to one half of a standard deviation. These effect sizes are comparable to those reported in meta-analyses of sex differences in temperament (Eaton & Enns, 1986; Else-Quest, Shibley Hyde, Goldsmith, & Van Hulle, 2006). More important to our twin analyses, MZ and DZ twins did not significantly differ in AL (home: $z = -.74$, $p = .46$; test: $z = -.68$, $p = .50$; and play: $z = .46$, $p = .65$).

AL Agreement Across Situations

Phenotypic correlations revealed substantial consistency in AL across the test and play situations ($r = .53$, $p < .001$). However, AL in the two laboratory situations were more modestly associated with AL at home (test: $r = .34$, $p < .001$ and play: $r = .33$, $p < .001$).

Twin Correlations

Twin intraclass correlations and cross-twin cross-situation correlations are presented in Table 2. For all three situations, the intraclass correlations for MZ twins exceeded those for DZ twins, suggesting genetic influences on AL. The very high intraclass correlations for both MZ and DZ twins in the home situation are particularly noteworthy and suggest that shared environmental influences also contribute to cotwin resemblance in this situation. Cross-twin cross-situation correlations for MZ twins were consistently higher that those for DZ twins, indicating that genetic factors contribute to the phenotypic correlation between AL across all three situations.

Model-Fitting Analyses

Univariate genetic analyses of AL in each situation (available from the first author) indicated that AL was significantly heritable in all three situations, suggesting that it is reasonable to inquire as to the extent to which there are cross-situational genetic effects (i.e., multivariate genetic analyses are warranted). Table 3 presents the fit statistics for the multivariate model-fitting analyses. As indicated by the nonsignificant

Table 1
Means (SD) for Activity Level by Sex and Zygosity in Home and Lab Situations

	Males		Females		Effect size	
Situation	MZ	DZ	MZ	DZ	Sex	Zygosity
Home	21.63 (2.49)	21.21 (2.45)	20.79 (2.45)	20.85 (2.10)	.24	.07
Test	24.71 (5.17)	24.37 (4.78)	22.35 (4.48)	22.08 (4.10)	.48	.05
Play	28.75 (7.24)	29.48 (6.53)	26.63 (6.81)	26.45 (6.86)	.38	−.05
n	146	182	142	154		

Note. Effect size estimated as Cohen's *d* expresses group differences in *SD* units. DZ = dzygotic twins; MZ = monozygotic twins.

Table 2
Twin Intraclass Correlations and Cross-Twin Cross-Situation Correlations (95% CI)

	Home		Test		Play	
Situation	MZ	DZ	MZ	DZ	MZ	DZ
Home	**.87 (.83 – .91)**	**.70 (.61 – .77)**				
Test	.30 (.14 – .44)	.18 (.03 – .32)	**.53 (.40 – .64)**	**.32 (.17 – .45)**		
Play	.31 (.16 – .45)	.19 (.04 – .33)	.49 (.36 – .61)	.27 (.13 – .41)	**.48 (.34 – .60)**	**.26 (.11 – .39)**

Note. Intraclass correlations are given in bold on the diagonal. Cross-twin correlations are given below the diagonal. DZ = dizygotic twins; MZ = monozygotic twins.

chi-square statistics for the overall fit of the model, both multivariate models adequately describe the data. Because the common pathway model is nested within the independent pathway model, it is possible to determine the relative fit of the two models by comparing the difference in chi-square statistics between the two models. As can be seen, imposing the more restrictive common pathway model to the data did not result in a significant decrement in the fit of the model ($\Delta\chi^2 = 1.589$, $df = 4$, $p = .811$). Moreover, its considerably lower Akaike's information criterion (difference = 6.411) provides substantial evidence for the common pathway model as the best-fitting overall model. We were able to further simplify the common pathway model by dropping all paths that were nonsignficant in the full model. The reduced model included additive genetic variance (A) on the latent AL phenotype; unique genetic, shared, and nonshared environmental effects on the home; and unique nonshared environmental effects on the test and play situations.

Figure 3 presents the path estimates from the full and reduced common pathway models. Post hoc power analyses indicated that we had more than 98% power to detect significant genetic covariance, but power to detect common shared environmental and nonshared environmental influences was lower, 23% and 39%, respectively. Therefore, although we present the results for both full and reduced models, to avoid artificial inflation of parameter estimates, we focus on the parameter estimates from the full model (i.e., in reduced models, nonsignificant shared environmental variance is absorbed into genetic variance). The path estimates are standardized partial regressions indicating the relative influence of the latent variables on the phenotypes. The square of these path coefficient estimates represent the proportion of variance for each measure accounted for by the latent variable. For example, the heritability of the latent AL phenotype is $.84^2$. The total genetic variance for AL in the home situation is the heritability of the latent AL factor times the variance in the AL measure that is due to the common factor (i.e., square of the factor loading) plus unique genetic variance (i.e., $[.84^2 \times .44^2] + .40^2 = .30$). Variance components from both models are presented in Table 4.

Table 3
Fit Statistics for Multivariate Models of Activity Level in the Home, Test, and Play Situations

	Overall fit of model[a]						Relative fit of model		
	−2LL	*df*	χ^2	*df*	*p*	AIC	$\Delta\chi^2$	*df*	*p*
Overall model (saturated −2LL = 9,967.52, 1,813)									
Independent pathway model	9,991.49	1,837	23.97	24	.46	−24.03			
Common pathway model	**9,993.08**	**1,841**	**25.56**	**28**	**.60**	**−30.44**	1.59[b]	4	.81
Simplification of best-fitting model									
No common C or E and no unique a or c on test or play	**9,997.04**	**1,847**	**29.52**	**34**	**.69**	**−38.49**	3.96[c]	6	.68

Note. Best-fitting models are indicated in bold. AIC = Akaike's information criterion; χ^2 = chi-square fit statistic; −2LL = log-likelihood statistic.
[a] Overall fit of the model is determined by the difference in −2LL of the model and that of a saturated model.
[b] χ^2 difference between the independent pathway model and the common pathway model.
[c] χ^2 difference between the full common pathway model and the reduced model.

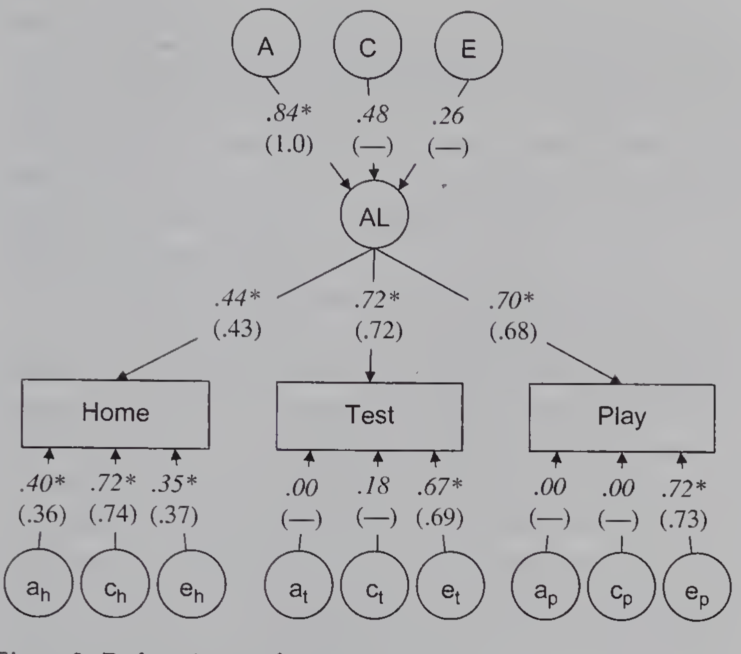


Figure 3. Path estimates from common pathway model.
Note. Estimates for the full model are given in italics. Asterisks indicate that the path was significant ($p < .05$) in the full model. Estimates for best-fitting reduced model are given in parentheses. Dashes indicate that the parameter was not included in the best-fitting reduced model. All paths in the reduced model are significant at $p < .05$.

Total variance components. AL was significantly heritable in all three situations, with genetic effects accounting for approximately one third of the total variance. As suggested by the twin intraclass correlations, AL in the home displayed significant and substantial shared environmental variance (i.e., more than 50% of the total variance). This pattern did not, however, emerge for AL in the laboratory. The environments that influenced AL in the test and play situations were of the nonshared variety.

Our multivariate model allowed us to test whether estimates of genetic and environmental variances significantly differed across situation. We fit models constraining either the genetic, shared environmental, or nonshared environmental variances to be equal across situations. These models were then compared to the full model. A significant change in chi-square statistic between models indicates that the constrained model provides a poorer fit to the data. It was possible to equate genetic variances ($\Delta\chi^2 = 0.636$, $df = 4$, $p = .96$), but neither the shared environmental variances ($\Delta\chi^2 = 20.418$, $df = 4$, $p < .001$) nor the nonshared environmental variances ($\Delta\chi^2 = 78.172$, $df = 4$, $p < .001$) could be equated across situations. Similar results emerged when genetic variances ($\Delta\chi^2 = 6.048$, $df = 4$, $p = .20$) or nonshared environmental variances ($\Delta\chi^2 = 88.672$, $df = 4$, $p < .001$) were equated in the reduced models. As indicated by the nonoverlapping confidence intervals, shared environmental variance was significantly higher and nonshared environmental variance significantly lower in the home situation as compared to AL in the lab.

Sources of covariance between situations. The significant factor loadings on the common latent AL factor indicate that the home, test, and play situations tap a common AL phenotype. This common AL factor explained 40% (i.e., $[.44^2 + .72^2 + .70^2]/3 = .40$) of the variance in our AL measures. Variation in the common AL phenotype is due to genetic influences (70%), neither shared nor nonshared environmental influences significantly influenced this common factor. In other words, covariance between situations (i.e., cross-situational consistency) arises because of common genetic influences. Genetic correlations indicating the extent to which genetic effects on AL in one situation correlate with genetic effects on AL in another situation, *independent of the heritability of each measure,* were substantial. The genetic effects on AL in the home correlated .68 (95% CI = 0.43 – 0.92) with AL in both laboratory situations. Similarly, the genetic correlation between AL in the test and in the play situations was 1.0 (95% CI = 0.78 – 1.0).

Cross-situational and situation-specific variance components. The genetic, shared, and nonshared environmental variances for each AL measure were decomposed into effects common across situations and those that are unique to a situation (Table 4). Approximately 50% of the genetic variance on AL in the home is situation specific (i.e., 16/30 = 53%). Neither AL in the test nor in the play situations displayed any unique genetic variance. Only AL in the home situation displayed significant shared environmental variance, accounting for more than 50% of the total variance on this measure; not surprising then was the finding of significant situation-specific shared environmental effects for the AL home measure. The only effects that were unique to the test and play situations were nonshared environmental influences, which include child-specific effects and measurement error.

Discussion

Do situations matter? There are two ways in which behavioral genetics research can approach this question. First, does the magnitude of genetic and environmental influence on AL differ across context or setting? A second, and perhaps less obvious, question involves the extent to which there is overlap

Table 4

Multivariate Estimates of Genetic and Environmental Variance (95% CI) From the Common Pathway Model

Full model

	Variance components		
	a^2	c^2	e^2
Latent AL phenotype	.70 (0.36–1.0)	.23 (0.00–0.53)	.07 (0.00–0.17)

	Loading on AL phenotype	Cross-situational variance			Situation-specific variance			Total variance		
		a^2	c^2	e^2	a^2	c^2	e^2	a^2	c^2	e^2
Home	.44 (.35–.53)	.14 (.07–.23)	.04 (.00–.13)	.01 (.00–.03)	.16 (.03–.31)	.52 (.38–.64)	.12 (.09–.17)	.30 (.17–.45)	.56 (.42–.68)	.14 (.11–.18)
Test	.72 (.65–.80)	.37 (.19–.57)	.12 (.00–.30)	.04 (.00–.10)	.00 (.00–.14)	.03 (.00–.12)	.44 (.35–.54)	.37 (.19–.59)	.15 (.00–.32)	.48 (.39–.57)
Play	.70 (.61–.76)	.34 (.17–.53)	.11 (.00–.27)	.03 (.00–.08)	.00 (.00–.13)	.00 (.00–.10)	.52 (.42–.61)	.34 (.17–.53)	.11 (.00–.27)	.55 (.45–.65)

Reduced model

	Variance components		
	a^2	c^2	e^2
Latent AL phenotype	1.0	—	—

	Loading on AL phenotype	Cross-situational variance			Situation-specific variance			Total variance		
		a^2	c^2	e^2	a^2	c^2	e^2	a^2	c^2	e^2
Home	.43 (.34–.51)	.18 (.12–.26)	—	—	.13 (.01–.28)	.55 (.41–.66)	.14 (.11–.18)	.31 (.19–.46)	.55 (.41–.66)	.14 (.11–.18)
Test	.72 (.66–.77)	.52 (.44–.60)	—	—	—	—	.48 (.40–.56)	.52 (.44–.60)	—	.48 (.40–.56)
Play	.68 (.62–.74)	.46 (.38–.55)	—	—	—	—	.54 (.45–.62)	.46 (.38–.55)	—	.54 (.45–.62)

Note. AL = activity level; a^2 = genetic variance; c^2 = shared environmental variance; e^2 = nonshared environmental variance.

between the genetic and the environmental effects on AL in the home and two laboratory situations. This latter approach addresses the issue of genetic and environmental influences to cross-situational consistency and context-specific effects. In the paragraphs below we will discuss situational effects from both perspectives.

As would be predicted from temperament theory, AL in all three situations was genetically influenced. There was little evidence of differential heritability across situation. Situational differences were, however, apparent in the nature and magnitude of environmental influences. AL in the home environment was substantially influenced by shared environmental factors (explaining 56% of the variance). This was not the case for AL in the laboratory test or play situations that displayed no significant shared environmental variance.

The finding of significant and substantial shared environmental variance on AL in the home is particularly interesting given that this finding does not emerge when parent ratings are used to assess AL (e.g., Goldsmith, Buss & Lemery, 1997; Neale & Stevenson, 1989; Saudino & Cherny, 2001b). Nonetheless, prior studies that have used mechanical measures to assess AL in the home environment have hinted at shared environmental influences. For example, Plomin and Foch (1980) used a weeklong pedometer measure to assess the AL of twins in middle childhood. Genetic influences on individual differences in AL were significant; however, the very high correlations between MZ and DZ twins suggested that shared environmental influences were substantial (MZ: r = .99, DZ: r = .94). Similarly, actometer-based studies of infants, toddlers, and adolescents have suggested possible shared environmental influences on AL, but these early studies were limited by very small sample sizes (Saudino & Eaton, 1991, 1995; Saudino, Thompson, & Gagne, 2000). The present findings provide the strongest evidence to date for shared environmental influences on mechanically assessed AL. The failure for parent ratings to display shared environmental influences likely arises from rater contrast biases, which magnify existing behavioral differences between cotwins. Indeed, when contrast biases have been included in behavioral genetic models of parent-rated AL, significant shared environmental effects have emerged (Saudino, Cherny, & Plomin, 2000; Saudino, McGuire, Reiss, Hetherington, & Plomin, 1995).

What shared environments might be contributing to variability in AL in the home? There are several possibilities. Family practices and schedules (e.g., meal times, bed times, family outings, and caretakers' tolerances for AL) are a likely source of common variance between cotwins. It is also possible that the increased shared environmental variance in the home situation reflects twin-specific environments. For example, the twins playing together and sharing toys may contribute to resemblance in AL. However, there are three lines of evidence that suggest that our finding of shared environmental influences on AL in the home is not simply a twin effect. First, MZ and DZ twins did not significantly differ in variances for home AL, suggesting that that shared environmental variance is not a result of cooperative sibling interaction effects. Second, for a subsample of families (n = 144), we asked parents to complete an activity diary listing the approximate amount of times that the twins engaged in common activities (e.g., storytime, preschool). In this subsample, the pattern of very high MZ and DZ twin resemblances (e.g., MZ: r = .83, DZ: r = .65) and significant shared environmental effects remained after residualizing home AL scores for time spent in common twin activities. Finally, a previous study of nontwin siblings also showed substantial resemblance for a 48-hr measure of actigraph AL (Saudino, Wertz, Gagne, & Chawla, 2004). The reported correlation for nontwin siblings in this earlier study (.58) does not significantly differ from our correlation for DZ twins and when compared to the MZ twin correlation also suggests shared environmental influences. This is particularly noteworthy because the nontwin siblings differed in age by approximately 3 years and because there were subtle methodological differences across the two studies. For example, children in the sibling study wore only two actigraphs (which reduces the reliability of the AL measure and would attenuate the sibling correlation) and the AL measure included time spent in the lab (which could also attenuate the nontwin sibling correlation). Thus, although we cannot definitively rule out twin-specific environments, it would appear that our finding of shared environmental influences in the home is not merely a twin-specific effect.

Whatever the source of these shared environmental effects, they do not have an enduring influence across situation. That is, although there are family-wide factors that influence AL in the home, they do not influence AL outside the home. As indicated previously, shared environmental influences did not significantly contribute to AL in either the laboratory test or the play situations. Moreover, our multivariate modeling clearly indicated that the shared environmental variance on AL in the home was situation specific.

The home and laboratory situations also differed in the magnitude of nonshared environmental influences operating on AL. In the home, nonshared environmental influences were modest, explaining less than 15% of the variance in AL. In contrast, approximately 50% of the variance in the two laboratory situations was due to nonshared environmental influences. In all cases, nonshared environmental influences were situation specific. Nonshared environmental variance comprises those environmental factors that are unique to each member of a twin pair and that make twins different from each other. This includes measurement error; however, given that the same devices were worn continuously across the 48-hr data collection period, differences in the magnitude of nonshared environmental variance across situation cannot be simply attributed to measurement error. One possible source of nonshared environmental variance in the two laboratory situations was the tester. Although testers followed standardized protocols in the test and play situations, tester personality, behavioral standards, and even mood may have influenced their interactions with the children tested. Additionally, twins were tested in the same situations on different days, which may also have contributed to nonshared environmental variance specific to each situation.

The fact that the magnitude of heritability of AL is similar across all three situations does not mean that the same genes operate in each situation. That is, it is possible that AL in each situation could represent a genetically influenced, but etiologically distinct, phenotype. This does not appear to be the case. Multivariate genetic analyses indicate that the covariation between AL in the home, test, and play situations was best described by a common latent factor, suggesting that to some extent, the same AL phenotype was assessed in all situations. Individual differences in this underlying AL phenotype were due only to genetic influences, indicating that what is common across all situations is genetic. Neither shared nor nonshared environmental influences contributed significantly to cross-situational consistency in AL.

Genetic correlations across situation indicate that approximately 68% of the genetic factors that influence AL in the home also influence AL in the laboratory test and play situations and that the genetic influences in the two laboratory situations are completely overlapping. This finding of substantial genetic overlap between AL across situations is accordant with prior research that finds that cross-situational consistency arises because of common genetic influences (i.e., Phillips & Matheny, 1997; Schmitz et al., 1996; Wood et al., 2007); however, our actigraph methodology and inclusion of the home situation adds more resolution. As with Wood et al., (2007) we found a genetic correlation of 1.0 between actigraph-assessed AL in two distinct laboratory situations (test and play in our study and test and break in the Wood et al., 2007, study). Phillips and Matheny (1997) report a genetic correlation of .59 between AL in the laboratory play and in the test situations, but as indicated earlier, they used different methods to assess AL across the two situations, which likely contributed to the lower genetic correlation. We suggest that when methodology is consistent across situations, AL within standard laboratory situations is not genetically distinct. This is not necessarily the case for AL beyond the laboratory. Although there was considerable overlap between AL in the home and in the laboratory, there were genetic influences unique to the home situation. Approximately 50% of the genetic variance on AL in the home was situation specific (i.e., independent of genetic variance on the common AL phenotype) indicating context-specific genetic effects for AL. Unlike prior research (Phillips & Matheny, 1997; Schmitz et al., 1996), our finding of contextual or situational-specific genetic effects is not confounded by method effects. Thus, genetic factors contribute to both cross-situational consistency *and* situational change. Environmental influences, however, contribute only to situational change.

The present results are consistent with temperament theory that predicts that individual differences in AL are genetically influenced and show some consistency across situations (Goldsmith et al., 1987). Our genetically sensitive design paired with actigraph methodology allowed us to examine the source of this consistency across situations and to evaluate sources of short-term situational change. The findings support both a cross-situational and a contextual view of the genetic underpinnings of temperament. Consistency in AL across situations is due to common genetic influences. However, even though genetic factors are solely responsible for cross-situational continuity, there is also evidence of situationally specific genetic effects on AL in the home. Our interpretation of these situation-specific effects is that different contexts provide different views of AL and thus engage different genetic influences on processes relevant to each context.

The limitations of this study should be acknowledged. Our sample of just more than 300 twin pairs is relatively small. Although we had ample power to detect significant genetic variance and covariance, our power to detect shared environmental effects on AL was more variable (ranging from approximately

20% in the lab to more than 90% in the home). For this reason, we highlight the results from the full model as reduced models that set nonsignificant nonzero parameters to zero can bias estimates. Similarly, very large samples are required to detect significant differences in heritabilities across measures; therefore, our sample does not provide a powerful test of differential heritability across situations. We would note, however, that the magnitudes of heritabilities across situations were very similar, ranging from .30 to .37 in the full model, and even in the reduced model where the range was greater, the 95% confidence intervals were overlapping. Despite the limitations imposed by our sample size, the present study makes a unique contribution to the developmental literature. Most studies of genetic influences on AL have relied on parent-rating measures or brief behavioral observations, few have considered situational effects. Our 48-hr actigraph measure of AL across three situations provides a novel and rigorous test of situational effects on AL from a behavioral genetic perspective.

In conclusion, although what is common among AL assessed in the three situations is largely genetic, situations do matter—provided that the situations are sufficiently different. Although the two standard lab conditions displayed no shared variance beyond the common genetic variance, the more diverse home environment revealed additional shared variance (both genetic and environmental) in AL. These results have important implications for molecular genetics research seeking to find specific genes associated with AL. The fact that to some extent, different genetic influences operate on AL across different situations means that researchers will need to consider the situation in which AL is assessed. Failures to replicate molecular genetic findings may be due to situational differences between studies. This would be especially true when comparing results across lab and nonlab situations. Molecular genetic studies of AL may benefit from a multisituation approach, which would allow identification of genes that are associated with the common AL phenotype (i.e., cross-situational genes). That is, given that the AL common factor was highly heritable, the search for cross-situational genes may increase the power of molecular genetic studies of AL.

References

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin, 101*, 213–232.

Barkley, R. A., & Ullman, D. G. (1975). A comparison of objective measures of activity and distractibility in hyperactive and nonhyperactive children. *Journal of Abnormal Child Psychology, 3*, 231–244.

Bayley, N. (1993). *Bayley Scales of Infant Development* (2nd ed.). San Antonio, TX: Psychological Corporation.

Braungart, J. M., Plomin, R., DeFries, J. C., & Fulker, D. W. (1992). Genetic influence on tester-rated infant temperament as assessed by Bayley's Infant Behavior Record: Nonadoptive and adoptive siblings and twins. *Developmental Psychology, 28*, 40–47.

Cherny, S. S., Fulker, D. W., Corley, R., Plomin, R., & DeFries, J. C. (1994). Continuity and change in infant shyness from 14 to 20 months. *Behavior Genetics, 24*, 365–380.

Cherny, S. S., Saudino, K. J., Fulker, D. W., Plomin, R. P., Corley, R., & DeFries, J. C. (2001). The development of shyness from 14 to 20 months: Shyness in context. In R. N. Emde & J. K Hewitt (Eds.), *The transition from infancy to early childhood: Genetic and environmental influences in the MacArthur Longitudinal Twin Study* (pp. 269–282). New York: Oxford University Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cyphers, L. H., Phillips, K., Fulker, D. W., & Mrazek, D. A. (1990). Twin temperament during the transition from infancy to early childhood. *American Academy of Child and Adolescent Psychiatry, 29*, 392–397.

Eaton, W. O. (1983). Measuring activity level with actometers: Reliability, validity, and arm length. *Child Development, 54*, 720–726.

Eaton, W. O., & Enns, L. R. (1986). Sex differences in human motor activity level. *Psychological Bulletin, 100*, 19–28.

Eaton, W. O., & Keats, J. G. (1982). Peer presence, stress, and sex differences in the motor activity levels of preschoolers. *Developmental Psychology, 18*, 534–540.

Else-Quest, N., Shibley Hyde, J., Goldsmith, H., & Van Hulle, C. (2006). Gender differences in temperament: A meta-analysis. *Psychological Bulletin, 132*, 33–72.

Fagot, B. I., & O'Brien, M. (1994). Activity level in young children: Cross-age stability, situational influences, correlates with temperament, and the perception of problem behaviors. *Merrill-Palmer Quarterly, 40*, 378–398.

Freedson, P. S., Sirard, J., Debold, N., Pate, R., Dowda, M., & Sallis, J. (1997, September). *Validity of two physical activity monitors in children and adolescents.* Paper presented at 19th Symposium of European Group of Pediatric Work Physiology, Moretonhamptonstead, UK.

Goldsmith, H. H., Buss, K. A., & Lemery, K. S. (1997). Toddler and childhood temperament: Expanded content, stronger genetic evidence, new evidence for the importance of environment. *Developmental Psychology, 33*, 891–905.

Goldsmith, H. H., Buss, A. H., Plomin, R., Rothbart, M. K., Thomas, A., Chess, S., et al. (1987). Roundtable: What is temperament? Four approaches. *Child Development, 58,* 505–529.

Goldsmith, H. H., Reilly, J., Lemery, K. S., Longley, S., & Prescott, A. (1995). *The laboratory temperament assessment battery-preschool version: Description of procedures.* Unpublished manuscript.

Hewitt, J. K., Silberg, J. L., Neale, M. C., Eaves, L. J., & Erikson, M. (1992). The analysis of parental ratings of children's behavior using LISREL. *Behavior Genetics, 22,* 292–317.

Hollingshead, A. B. (1975). *The four-factor index of social status.* Unpublished manuscript, Yale University, New Haven, CT.

Janz, K. F. (1994). Validation of the CSA accelerometer for assessing children's physical activity. *Medicine and Science in Sports and Exercise, 26,* 369–375.

Kendler, K. S., Heath, A. C., Martin, N. G., & Eaves, L. J. (1987). Symptoms of anxiety and symptoms of depression: Same genes, different environments? *Archives General Psychiatry, 44,* 451–457.

Liang, K. Y., & Zeger S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika, 73,* 13–22.

Mangelsdorf, S. C., Schoppe, S. J., & Burr, H. (2000). The meaning of parental reports: A contextual approach to the study of temperament and behavior problems in childhood. In V. J Molfese & D. L Molfese (Eds.), *Temperament and personality development across the lifespan* (pp. 1–32). Mahwah, NJ: Earlbaum.

Matheny, A. P., Jr. (1980). Bayley's Infant Behavior Record: Behavioral components and twin analysis. *Child Development, 51,* 1157–1167.

Matheny, A. P., Jr. (1983). A longitudinal twin study of stability of components from Bayley's Infant Behavior Record. *Child Development, 54,* 356–360.

McArdle, J. J., & Goldsmith, H. H. (1990). Alternative common-factor models for multivariate biometric analyses. *Behavior Genetics, 20,* 569–608.

McGue, M., & Bouchard, T. J. (1984). Adjustment of twin data for the effects of age and sex. *Behavior Genetics, 14,* 325–343.

Melanson, E., & Freedson, P. S. (1995). Validity of the Computer Science and Applications, Inc. activity monitor. *Medicine & Science in Sports & Exercise, 72,* 934–940.

Mini Mitter Company. (2005). *Actical physical monitoring system instruction manual.* Bend, OR: Author.

Neale, M., Boker, S. M., Xie, G., & Maes, H. H. (2006). *Mx: Statistical modeling* (7th ed.). Richmond, VA: Virginia Commonwealth University, Department of Psychiatry.

Neale, M. C., & Stevenson, J. (1989). Rater bias in the EASI temperament scales: A twin study. *Journal of Personality and Social Psychology, 56,* 446–455.

Phillips, K., & Matheny, A. P. (1997). Evidence for genetic influence on both cross-situation and situation-specific components of behavior. *Journal of Personality and Social Psychology, 73,* 129–138.

Pinto, L. P., & Tryon, W. W. (1996). Activity measurements support dimensional assessment. *Behavior Modification, 20,* 243–258.

Plomin, R., DeFries, J. C., McClearn, G. E., & McGuffin, P. (2001). *Behavioral genetics* (4th ed.). New York: Worth.

Plomin, R., & Foch, T. T. (1980). A twin study of objectively assessed personality in childhood. *Journal of Personality and Social Psychology, 39,* 680–688.

Plomin, R., & Nesselroade, J. R. (1990). Behavioral genetics and personality change. *Journal of Personality, 58,* 191–220.

Price, T. S., Freeman, B., Craig, I., Ebersole, L., & Plomin, R. (2000). Infant zygosity can be assigned by parent questionnaire data. *Twin Research, 3,* 129–133.

Saudino, K. J., & Cherny, S. S. (2001a). Sources of continuity and change in observed temperament. In R. N. Emde & J. K Hewitt (Eds.), *The transition from infancy to early childhood: Genetic and environmental influences in the MacArthur Longitudinal Twin Study* (pp. 89–110). New York: Oxford University Press.

Saudino, K. J., & Cherny, S. S. (2001b). Parent ratings of temperament in twins. In R. N Emde & J. K. Hewitt (Eds.), *The transition from infancy to early childhood: Genetic and environmental influences in the MacArthur Longitudinal Twin Study* (pp. 73–88). New York: Oxford University Press.

Saudino, K. J., Cherny, S. S., & Plomin, R. (2000). Parent ratings of temperament in twins: Explaining the "too low" DZ correlations. *Twin Research, 3,* 224–233.

Saudino, K. J., & Eaton, W. O. (1991). Infant temperament and genetics: An objective twin study of motor activity level. *Child Development, 62,* 1167–1174.

Saudino, K. J., & Eaton, W. O. (1995). Continuity and change in objectively assessed temperament: A longitudinal twin study of activity level. *British Journal of Developmental Psychology, 13,* 81–95.

Saudino, K. J., McGuire, S., Reiss, D., Hetherington, E. M., & Plomin, R. (1995). Parent ratings of EAS temperaments in twins, full siblings, half siblings and step siblings. *Journal of Personality and Social Psychology, 68,* 723–733.

Saudino, K. J., Plomin, R., & DeFries, J. C. (1996). Tester-rated temperament at 14, 20, and 24 months: Environmental change and genetic continuity. *British Journal of Developmental Psychology, 14,* 129–144.

Saudino, K. J., Thompson, L. A., & Gagne, J. R. (2000). Genetic influences on activity level in adolescence: Do the same genetic effects operate on parent ratings, self ratings and mechanical measures? [Abstract]. *Behavior Genetics, 30,* 417–418.

Saudino, K. J., Wertz, A., Gagne, J. R., & Chawla, S. (2004). Night and day: Are siblings as different in temperament as parents say they are? *Journal of Personality and Social Psychology, 87,* 698–706.

Schaughency, E. C., & Fagot, B. I. (1993). The prediction of adjustment at age 7 from activity level at age 5. *Journal of Abnormal Child Psychology, 21,* 29–50.

Schmitz, S., Saudino, K. J., Plomin, R., Fulker, D. W., & DeFries, J. C. (1996). Genetic and environmental influences on temperament in middle childhood: Analyses of teacher and tester ratings. *Child Development*, *67*, 409–422.

Stevenson, J., & Fielding, J. (1985). Ratings of temperament in families of young twins. *British Journal of Developmental Psychology*, *3*, 143–152.

Wood, A., Saudino, K. J., Rogers, H., Campbell, D., Asherson, P., & Kuntsi, J. (2007). Genetic influences on mechanically-assessed activity level in children. *Journal of Child Psychology and Psychiatry*, *48*, 695–702.

Zeger, S. L., & Liang, K. Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, *42*, 121–130.

Child Development, July/August 2008, Volume 79, Number 4, Pages 944–956

Stability and Change in Ethnic Labeling Among Adolescents From Asian and Latin American Immigrant Families

Andrew J. Fuligni
University of California, Los Angeles

Lisa Kiang
Wake Forest University

Melissa R. Witkow
Willamette University

Oscar Baldelomar
University of California, Los Angeles

An important question for the acculturation of adolescents from immigrant families is whether they retain ethnic labels that refer to their national origin (e.g., Mexican, Chinese) or adopt labels that are dominant in American society (e.g., Latino, Asian American, American). Approximately 380 adolescents from Asian and Latin American immigrant families selected ethnic labels during each of the 4 years of high school (age span = 14.87 – 17.82 years). Results indicated no normative trend either toward or away from identifying most closely with pan-ethnic or American ethnic labels. Significant numbers of adolescents changed their ethnic labels from year to year, however and these changes were associated with fluctuations in adolescents' ethnic affirmation and exploration and proficiency in their heritage languages.

Given the social importance and functional significance of ethnicity in American society, children are virtually required to figure out the ethnic and social categories into which they are placed. This task can be particularly challenging for children and adolescents from immigrant families who are less familiar with the ethnic categories and labels that are used in the United States. An important question for the acculturation of adolescents from immigrant families, therefore, is whether they retain ethnic labels that refer to their national origin (e.g., Mexican, Chinese) or move toward the adoption of labels that are dominant in American society (e.g., Latino, Asian American, American). Given that immigrant families comprise the large majority of those with Asian and Latin American backgrounds and that these are the two fastest rising ethnic groups in the United States, the answer to this question will have implications for the nature of ethnic categories and the development of ethnic identity in the broader society.

Ethnic Labeling Among Immigrants

An important aspect of identity development during the adolescent years is the process by which teenagers explore how they fit into existing social categories and groups (Ruble et al., 2004). Ethnicity is a meaningful and salient social category within the United States, differentiating people across an array of social, economic, and behavioral indicators. The process of ethnic identity development is particularly significant for those from ethnic minority and immigrant groups because their minority and immigrant status make their ethnicity more salient in the eyes of American society. A large body of research has documented the significance of ethnic identity for critical aspects of development among ethnic minority and immigrant adolescents, such as self-esteem, academic motivation, and behavioral adjustment (Phinney, 1990). Much of this work has focused on how adolescents explore or value their ethnic group membership, however, and much less is known about how teenagers come to adopt the specific ethnic labels that they use to describe themselves. Yet, the limited work on ethnic labeling has suggested that it can be consequential for adolescent development. For example, both Waters (1999) and Matute-Bianchi (1991) noted how ethnic labeling can channel adolescents from immigrant families into different peer groups, with those who identify as "Mexican" or "Jamaican" having very different social worlds than those who identify as "Chicano" or "African American." In addition, some studies have noted how these labels

---

Support for this study was provided by the Russell Sage Foundation. The authors also thank the participating students and schools for their assistance.

Correspondence concerning this article should be addressed to Andrew J. Fuligni, Center for Culture and Health, University of California, Los Angeles, 760 Westwood Plaza, Box 62, Los Angeles, CA 90024. Electronic mail may be sent to afuligni@ucla.edu.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7904-0010

also can be associated with achievement orientations and school success, with immigrant adolescents who identify with national origin labels such as "Mexican" evidencing higher levels of school success (Matute-Bianchi, 1991; Rumbaut, 1994).

Although recent research has demonstrated how adolescents can employ multiple ethnic labels to describe themselves in different contexts and settings (Harris & Sim, 2002; Moje & Martinez, 2007), an important question for adolescents from immigrant families is the extent to which they will identify with ethnic labels and categories that are distinctive to the United States and about which they may be less familiar. For example, "Latino" and "Hispanic" are pan-ethnic labels that are commonly used to refer to individuals of Latin American origin in the United States but that are absent in Latin American societies (Portes & Rumbaut, 2001). Similarly, although the terms "Asian" and "White" may not be entirely new to immigrants, their social significance is particularly American because they place adolescents within the major ethnic and racial categories used by the government and most social institutions (i.e., Black, White, Asian, Hispanic, and Native American). Another convention unique to the United States is to use the term "American" by itself or to pair national or pan-ethnic labels with the term "American" to create hyphenated labels such as "Mexican-American" and "Asian-American."

A few cross-sectional studies have indicated that adolescents from immigrant families, particularly adolescents who were born outside of the United States themselves, are somewhat resistant to adopt pan-ethnic and American ethnic labels. A survey of eighth- and ninth-grade students in Florida and Southern California reported a greater tendency to choose national origin labels such as "Mexican" and "Chinese" among the first-generation (i.e., foreign born) adolescents as compared to their second-generation (i.e., American born) peers, who were more likely to use pan-ethnic and American or hyphenated-American labels (Rumbaut, 1994). Ethnographic studies have suggested that some foreign-born adolescents, such as Black immigrants from the West Indies and immigrants from Mexico, seek to retain their national origin labels as a way to avoid the negative stereotypes associated with labels such as Black, Hispanic, or Latino (Matute-Bianchi, 1991; Waters, 1999). The pattern of first-generation adolescents preferring national origin labels and second-generation adolescents choosing pan-ethnic, American, and hyphenated-American labels has been observed in other samples (Buriel & Cardoza, 1993; Doan & Stephan, 2006), including ninth-grade students who participated in the first wave of the longitudinal study that is reported in the present article (Fuligni, Witkow, & Garcia, 2005).

Generational differences such as these are often mistakenly interpreted as meaning that spending more time in the United States will produce changes within immigrants themselves. But generational differences may not translate into changes within individuals across time and development due to the numerous confounding factors associated with generational status such as birthplace and the unique historical context within which each generation entered the country (Fuligni, 2001). It is important to determine whether ethnic labels become more Americanized within children from immigrant families across time and development to determine whether indeed more time in the United States leads to an assimilation of the society's dominant ethnic categories. As a key time of ethnic identity formation, the adolescent years are an important developmental period to study this process.

Longitudinal examinations of changes in ethnic labeling are rare and there are competing predictions about what might be observed. On the one hand, the adolescent period could be a period of significant change in ethnic labeling among those from immigrant families as they become more aware of the existence and the social significance of the ethnic categories used in American society (Doan & Stephan, 2006). The continual need in secondary school to complete official forms that require identification with a pan-ethnic group, such as Asian or Hispanic, a general desire to belong to and fit in with normative groups and social categories, and an increasing awareness of the ethnic landscape of American society could lead those from immigrant families to adopt pan-ethnic, American, or hyphenated-American ethnic labels as they progress through high school. Like most adolescents in the United States, children from immigrant families engage in the normative process of emotionally distancing themselves somewhat from their parents during adolescence (Fuligni, 1998), and this process could weaken adolescents' connection to their families' national origin and result in a movement toward other ethnic labels.

On the other hand, there may be stability in the ethnic labels chosen by adolescents from immigrant families over the high school years. Some observers have suggested that birthplace is a primary criterion by which individuals may determine their ethnic categories (Rumbaut, 1994). First-generation, foreign-born adolescents may therefore retain the national origin labels that they have been shown to prefer in previous cross-sectional studies because they believe that they will always be primarily "Mexican" or "Chinese" because of their birth in Mexico or China. Similarly, second-generation adolescents may feel unable to adopt such a national origin label because they were not born in those countries. Instead, those of the second generation may identify with labels such as Latino or Asian American throughout the adolescent years because they were born in the United States. Another possibility is that adolescents from immigrant families move toward the incorporation of national original labels even more, as their process of identity development leads them to explore their family's national and cultural origins more closely.

The Current Study

We sought to test these competing hypotheses in a longitudinal study of adolescents from Asian and Latin American immigrant families who were followed yearly during the 4 years of high school. First, we examined whether there was a normative change toward the inclusion of pan-ethnic and American (either hyphenated or alone) terms in ethnic labels as adolescents progressed through the high school years. Second, we assessed whether changes in the use of pan-ethnic and American terms varied according to the adolescents' ethnic background, generation, and gender. It is possible that adolescents from Asian immigrant families would be more likely to move in the direction of the adoption of pan-ethnic or American terms. Because of the greater economic integration of their parents and their own academic success at school, adolescents from Asian immigrant families may identify with American society more quickly than those from Latin American immigrant families (Portes & Zhou, 1993). In addition, individual Asian immigrant groups are relatively smaller in size than those from Latin America, which may lead toward greater aggregation across specific ethnic groups into a larger pan-ethnic identification. Indeed, in perhaps the only longitudinal examination of changes in ethnic labeling, Portes and Rumbaut (2000) reported that adolescents from Chinese immigrant families were more likely to move toward the adoption of a pan-ethnic label between middle school and the end of high school, whereas those from Mexican immigrant families moved away from pan-ethnic labels and toward the national origin label of "Mexican."

In terms of generational status, it is possible that there would be greater change among the foreign-born adolescents, particularly among those who come to the United States at older ages, because they start out with the endorsement of the national origin label, thereby giving them more room to move toward the inclusion of pan-ethnic and American terms as they spend more time in the United States. Second-generation adolescents, by virtue of being born in the United States, may feel that they have fewer ethnic options available to them and must conform to the ethnic categories that are dominant in American society. Finally, it is difficult to predict whether there would be gender differences in the changes in adolescents' choice of ethnic labels. Our earlier analyses of ninth-grade students who participated in the longitudinal study reported in the present article revealed no gender differences in adolescents' choice of their most descriptive ethnic labels, and gender differences were inconsistent in the Portes and Rumbaut (2001) study discussed earlier. Nevertheless, we explored potential gender differences given some previous evidence suggesting that girls in ethnic minority families were more likely to be socialized by their parents in their ethnic and cultural background (Bowman & Howard, 1985), which could lead to a lower likelihood of girls moving toward the inclusion of pan-ethnic and American terms in their ethnic labels.

In addition to examining variations in changes *between* adolescents, we analyzed changes *within* adolescents over time as a function of simultaneous changes in their family cohesion, ethnic exploration and affirmation, and proficiency in their families' heritage languages (e.g., Chinese, Spanish). Such analyses allowed us to get closer to the processes by which changes in ethnic labeling take place within adolescents by looking at how these changes are associated with changes in other significant aspects of their social and psychological development. We selected these factors because research on adolescents from immigrant families suggests that ethnic labeling may be part of a larger acculturation process that involves changes in family relationships, ethnic exploration and attachment, and language use (Phinney, 2003; Portes & Rumbaut, 2001; Tseng & Fuligni, 2000). These factors often are associated with one another on an individual differences basis, suggesting that they are part of a general process by which adolescents from immigrant families negotiate the potential differences between their cultural background and American society. As such, we believed that they might be candidates for predictors of changes in ethnic labeling *within* individual adolescents over the high school years.

In terms of family cohesion, adolescence is a period of fluctuation in parent–child relationships, and these relationships are thought to be critical factors in the retention of traditional cultural orientations in immigrant families (Gonzales, Deardorff, Formoso, Barr, & Barrera, 2006). Several observers have suggested that acculturation stress may occur between immigrant parents and their children because of the different rates at which they adapt to American society (Zhou, 1997). Therefore, it is possible that as adolescents become more emotionally distant from their parents over the high school years, their ties to their families' national and cultural origins may weaken. As a result, the adolescents may move away from ethnic labels that incorporate their national origin and toward labels that include pan-ethnic or American terms.

The strength of adolescents' ethnic identity refers to the extent to which adolescents identify with and feel attached to their ethnic group (Ashmore, Deaux, & McLaughlin-Volpe, 2004; Phinney, 1990). Two predictions could be made about the links between the strength of adolescents' ethnic identity and their choice of ethnic labels. On the one hand, stronger ethnic identity could be associated with a greater adoption of pan-ethnic and American ethnic terms. This is because a stronger ethnic identity would reflect a greater exploration of ethnic identity in the context of an American society that uses these terms to demarcate the dominant ethnic categories. In addition, stronger ethnic identity could result in a perception of common fate among members of different Asian or Latin American ethnicities, resulting in greater pan-ethnic identification (Lien, Conway, & Wong, 2003).

On the other hand, higher levels of ethnic identity could be associated with the retention of national labels because the identity exploration of adolescents from immigrant families would require them to consider their national heritage. In addition, it could simply be harder to maintain a national label in American society, thereby requiring the identity work that would be reflected in higher levels of ethnic identity. In support of the latter hypothesis, earlier analyses of ninth-grade students who participated in the first wave of the longitudinal study that is reported in the present article suggested that ethnic identity was stronger among adolescents who primarily chose a national origin label (Fuligni et al., 2005).

Our expectation for the links between changes in adolescents' proficiency in their heritage language and their ethnic labels was straightforward. Given the meaning of language use and proficiency for ethnic and cultural identity that has been observed previously among immigrant families and adolescents (e.g., Phinney, Romero, Nava, & Huang, 2001), lower levels of adolescents' heritage language proficiency across the high school years were expected to be associated with a less use of the national origin label and more use of pan-ethnic and American terms.

Method

Sample

Beginning in 9th grade and continuing yearly through 12th grade, students from three public high schools in the Los Angeles area were recruited for participation in a longitudinal study. The first school was populated predominantly by students who came from families with Latin American and Asian backgrounds with lower-middle to middle-class educational and occupational statuses. This school tended to be in the lower middle to middle range of the achievement distribution of schools within the state of California (California Department of Education, 2006). The second school possessed average levels of achievement and consisted mainly of students from Latin American and European backgrounds whose families had lower-middle to middle-class backgrounds. Finally, the third school mainly consisted of students from families with Asian and European backgrounds who were middle- to upper-middle-class in terms of parental education and occupation. The third school tended to have above average achievement levels. No single ethnicity dominated any of these schools; rather, the two largest ethnic groups each comprised 30%–50% of the total population in each school (California Department of Education, 2006).

In two of the schools, the entire ninth grade was invited to participate during the 1st year of the study. At these schools, the same process continued in subsequent years, with all students in the correct grade being invited to participate. In the third school, approximately half the ninth graders were invited to participate because the large size of the school did not make it feasible to recruit all the students. In this school, only students who had participated in ninth grade were followed in subsequent years. At all three schools, students who had participated in earlier years but were no longer enrolled in the school were contacted and invited to participate by mail in subsequent years.

The sample used in the present analyses was the 384 participants from Asian and Latin American immigrant families who had completed self-report questionnaires for at least 2 of the 4 years of the study, where an immigrant family was defined as one in which at least one parent was foreign born. The mean ages of the sample ranged from 14.87 years (*SD* =

0.39) at the 9th grade to 17.82 years ($SD = 0.38$) at the 12th grade. The sample was evenly split by sex (187 male and 197 female). The majority of the 228 participants from Asian immigrant families were from Chinese backgrounds (76.8%), and the majority of the 156 participants from Latin American immigrant families had Mexican backgrounds (86.5%). Of the Asian participants, 78 were of the first generation (i.e., the students were foreign-born themselves) and 150 were of the second generation (i.e., the students were born in the United States). Of the participants from Latin American families, 35 were of the first generation and 121 were of the second generation. The average ages at which first-generation adolescents came to the United States were 7.88 and 4.07 years (SDs = 4.15 and 3.44) for those from Asian and Latin American backgrounds, respectively.

Adolescents from the two immigrant groups differed in terms of parental education such that the Asian immigrant parents were more likely to have graduated from high school and attended some college than the parents from Latin America, $F(1, 371) = 28.58, p < .001$. There were no differences in parental education according to the adolescents' generational status, $F(1, 371) = .89$, *ns*, and the relationship between ethnicity and education did not vary according to generational status, $F(1, 371) = .07$, *ns*.

Measures

Participants completed the following measures in self-report questionnaires during each year of the study. Questionnaires were available to students in English, Chinese, and Spanish and fewer than 8 participants chose to complete the questionnaires in a language other than English during any single year.

Ethnic labeling. Adolescents were presented with an alphabetical list of ethnic labels that is shown in Appendix. The list was developed based on discussions with ethnically diverse adolescents, discussions with Mexican and Chinese undergraduate research staff members who attended high school in the Los Angeles area, and prior research on the type of labels used by adolescents from these different groups (e.g., Matute-Bianchi, 1991; Rumbaut, 1994). Adolescents were asked to indicate all the labels that they felt applied to them and to add any others that applied to them but that were not on the list (e.g., mixed-race labels, other ethnic labels). Additionally, students were asked to indicate the ethnic label that they believed best described them. Adolescents were allowed to report more than one label for the last question about the best descriptive label if they felt that it was most accurate for their own ethnic identity, but this was done by less than 3% of the sample during any given year.

Each year, labels were coded as to whether a pan-ethnic term (e.g., Asian, Latino) or an American term, either by itself or hyphenated (e.g., American, Mexican-American), was included in the label. In addition, labels were coded as to whether they referred to solely national origin (e.g., Mexican, Chinese). The vast majority of adolescents included pan-ethnic/American labels (98.7%) and national origin labels (89.6%) in their total list of chosen labels during at least 1 year of the study, providing support to the idea that both types of labels are used by adolescents during the high school years. Our analyses focused on the labels with which adolescents identified most strongly, which were those that adolescents reported described them best. For these analyses, the most descriptive label was coded as to whether it included a pan-ethnic or an American term, either by itself or hyphenated (0 = *no*, 1 = *yes*). Similar findings to those reported in the Results section were found when choice of pan-ethnic and American labels were each analyzed separately. Of the 3% of adolescents who reported multiple most descriptive labels, all but two adolescents reported labels that could both be classified into the same categories described earlier (e.g., pan-ethnic and national origin). The two participants who listed multiple most descriptive labels that fell into different categories were not included in the present analyses. The specific labels chosen by adolescents as most descriptive each year are presented in Table 1.

Closeness to parents. Adolescents completed a subset of the Family Adaptation and Cohesion Evaluation Scales II inventory separately for each parent (Olson, Sprenkle, & Russell, 1979). Questions were the same for mothers and fathers and participants responded to each on a 5-point scale (1 = *almost never*, 5 = *almost always*). The measure for each parent included 10 items such as "My mother [father] and I do things together," "My mother [father] and I are supportive of each other during difficult times," and "My mother [father] and I feel very close to each other." An average was computed for each participant across both parents. Internal consistencies were similarly high for both ethnic groups across the 4 years of the study (αs = .87 – .92). Scores were recoded on a 0–4 scale for the analyses presented in this article.

Strength of ethnic identity. Adolescents completed two subscales of the Multigroup Ethnic Identity Measure (Phinney, 1992) in reference to the most descriptive ethnic label that they chose each year.

Table 1
Frequencies of Most Descriptive Labels

	Latin American				Asian			
Label	9th	10th	11th	12th	9th	10th	11th	12th
American	1.7	3.6	1.6	3.9	1.1	2.0		2.5
Asian					7.9	5.9	11.6	4.5
Asian American					23.7	20.0	23.6	18.3
Asian Indian							0.5	
Australian					0.6			
Brazilian	1.7	1.4	0.8	1.6				
Burmese								0.5
Cantonese						0.5		
Central American		0.7	0.8					
Chicano	2.6	3.6	6.3	3.1				
Chinese					16.4	20.0	14.6	14.9
Chinese American					19.8	22.0	17.6	26.7
Filipino					3.4	3.9	3.5	3.5
Guatemalan	3.4	3.6	3.2	4.7				
Hispanic	4.3	13.7	9.5	12.4				
Hispanic American	12.8	5.0	4.8	4.7				
Honduran	0.9							
Indonesian								0.5
Japanese					1.1	1.0	1.0	1.0
Japanese American					3.4	3.4	2.5	2.5
Korean					2.3	2.4	2.5	2.5
Korean American					4.5	3.9	5.5	5.0
Latino	10.3	13.7	8.7	10.1				
Latino American	5.1	5.0	4.8	3.9				
Mexican	18.8	21.6	23.0	16.3				
Mexican American	33.3	25.9	31.0	35.7				
Nicaraguan American	0.9		0.8	0.8				
Pacific Islander								0.5
Peruvian		0.7	0.8					
Puerto Rican	0.9							
Salvadoran	0.9	0.7	3.2	1.6				
Salvadoran American	0.9	0.7	0.8	1.6				
Spanish	0.9							
Taiwanese					2.3	2.0	4.0	2.0
Taiwanese American					2.3	2.0	2.0	3.5
Thai					1.1	0.5	0.5	0.5
Uruguayan	0.9							
Vietnamese					3.4	5.4	3.5	3.5
Vietnamese American					6.8	5.4	7.0	7.9

Note. Figures refer to the percentage of adolescents within each ethnic group at each grade level who chose each label as their most descriptive label. Any label that included the term "American," by itself or in combination with another term, was coded as including an American term except for "Central American." Any label that included "Asian," "Chicano/a," "Latino," and "Hispanic," either alone or in combination with another term, was coded as including a pan-ethnic term.

The Affirmation and Belonging subscale consists of seven items and assesses ethnic pride, feeling good and happy about one's ethnicity, and feelings of belonging and attachment to one's ethnic group. Sample items include "I am happy that I am a member of the group I belong to," "I have a lot of pride in my ethnic group" and "I feel good about my cultural or ethnic background." The Ethnic Identity Search subscale consists of five items and measures individuals' exploration of and commitment to their ethnic group. Sample items include "I have spent time trying to find out more about my ethnic group, such as its history, traditions, and customs," "I am active in organizations or social groups that include mostly members of my own ethnic group," and "In order to learn more about my ethnic background, I have often talked to other people about my ethnic group." Participants responded to each item on a 5-point scale (1 = *strongly disagree*, 5 = *strongly agree*). Internal consistencies were similarly high for both ethnic groups across the 4 years of the study for both measures (affirmation: αs = .89 – .91, search: αs = .68 – .78). Scores were recoded on a 0 – 4 scale for the analyses presented in this article.

Heritage Language Proficiency

Using a standard approach to measuring language proficiency in large-scale questionnaire studies that prohibit the direct testing of language proficiency (e.g., Phinney et al., 2001; Portes & Rumbaut, 2001), adolescents were asked to report all the languages that were spoken in their home. For any non-English language spoken in the home (e.g., Chinese, Spanish), participants were asked to rate how well they speak, understand, read, and write that language. Participants responded to each item on a 5-point scale (1 = *not very well*, 5 = *very well*), and responses to these multiple items of proficiency were averaged. Participants received a 0 on this scale if they indicated that no non-English languages are spoken in their home, resulting in a scale that ranged from 0 to 5. Internal consistencies were similarly high for both ethnic groups across the 4 years of the study (αs = .86 – .89).

Results

Participation Analyses

Initial analyses were conducted to examine differences between adolescents according to the degree of their participation in the study. Although this sample could include those who participated only 2 years, most participated all 4 years of the study (*M* = 3.52 years, *SD* = 0.67). Because participants entered the study at different years, a variable was created to indicate the percentage of possible years each participant took part in the study. For example, a participant who began the study in 9th grade had four possible years, whereas a participant who began the study in

10th grade only had three possible years. On average, participants took part in the study in 94% ($SD = 0.14$) of their possible years. There were no differences in the degree of participation as a function of gender, ethnicity, or generation. To determine if there were differences in any of the time varying variables (i.e., ethnic labeling, closeness to parents, strength of ethnic identity, and heritage language proficiency) as a function of participation, hierarchical linear models (HLM; Bryk & Raudenbush, 1992) were estimated using the following equations:

$$\text{Ethnic labeling, etc.}_{ij} = b_{0j} + e_{ij} \qquad (1)$$

$$b_{0j} = c_{00} + c_{01}(\text{participation}) + u_{0j} \qquad (2)$$

where Equation 1 represents adolescents' scores on the time varying variables (i.e., ethnic labeling, closeness to parents, etc.) across the years of their participation in the study and Equation 2 represents the prediction of the adolescents' average scores across their years of participation as a function of their degree of participation (i.e., the percentage of possible years that they took part in the study). There were no differences in any of the variables in the study based on degree of participation except for heritage language proficiency, where adolescents who participated in more years of the study had lower average levels of heritage language proficiency ($b = -1.40$, $p < .01$).

For this analysis and all subsequent analyses described subsequently, additional hierarchical generalized linear models (HGLM) were tested because the outcome variable had only two possible values. Except in the one minor case noted subsequently, the substantive results remained unchanged. For ease of presentation and interpretation, results from the HLM are presented throughout.

Normative Changes in Ethnic Labeling

The first goal of this study was to examine change over time in adolescents' inclusion of a pan-ethnic or an American term in their most descriptive ethnic label. The statistical model that was estimated was as follows:

$$\text{Panethnic or American term}_{ij} = b_{0j} + b_{1j}(\text{year}) + e_{ij} \qquad (3)$$

$$b_{0j} = c_{00} + c_{01}(\text{gender}) + c_{02}(\text{ethnicity}) + c_{03}(\text{generation}) + u_{0j} \qquad (4)$$

$$b_{1j} = c_{10} + c_{11}(\text{gender}) + c_{12}(\text{ethnicity}) + c_{13}(\text{generation}) + u_{1j} \qquad (5)$$

As shown in Equation 3, adolescents' inclusion of a pan-ethnic or an American term (0 = *no*, 1 = *yes*) on a particular year (i) for a particular individual (j) was modeled as a function of the average amount of inclusion of the term by the individual (b_{0j}) and the year of the study (b_{1j}). Year was coded such that *Year 1* = 0, *Year 2* = 1, *Year 3* = 2, and *Year 4* = 3. Equations 4 and 5 show how both the average amount of inclusion and the effect of the year of the study were modeled as a function of the adolescents' gender, ethnicity, and generational status. The Level 2 variables were effects coded such that gender was coded as *males* = −1 and *females* = 1, ethnicity was coded as *Latino* = −1 and *Chinese* = 1, and generation was coded as *first generation* = −1 and *second generation* = 1.

Results indicated no normative developmental changes in adolescents' inclusion of a pan-ethnic or an American term in their most descriptive ethnic labels. On average, second-generation adolescents were more likely to choose a pan-ethnic or an American term (Table 2). But as shown in Table 2 and Figure 1, there were no significant changes in ethnic labeling over time ($b = -.02$, *ns*). Further, there was no significant variation in the effect of time on labeling (*SD* of estimate = 0.05, *ns*) and the amount of change did not significantly vary across adolescents' gender, ethnicity, and generation.

An additional analysis was conducted with first-generation participants to test whether amount of change varied according to age of immigration. For

Table 2
Hierarchical Linear Models Predicting Change Over Time in the Inclusion of a Pan-Ethnic or an American Term in Adolescents' Most Descriptive Ethnic Labels

	Pan-ethnic or American term, *b* (*SE*)
Intercept	.58 (0.03)***
Gender	−.01 (0.02)
Ethnicity	−.01 (0.02)
Generation	.11 (0.03)***
Year	−.02 (0.01)
Gender	.00 (0.01)
Ethnicity	.01 (0.01)
Generation	−.01 (0.01)

Note. Gender was coded as *boys* = −1, *girls* = 1; ethnicity was coded as *Latino* = −1, *Asian* = 1; and generation was coded as *first generation* = −1, *second generation* = 1. All predictors were uncentered.
$^{***}p < .001$.

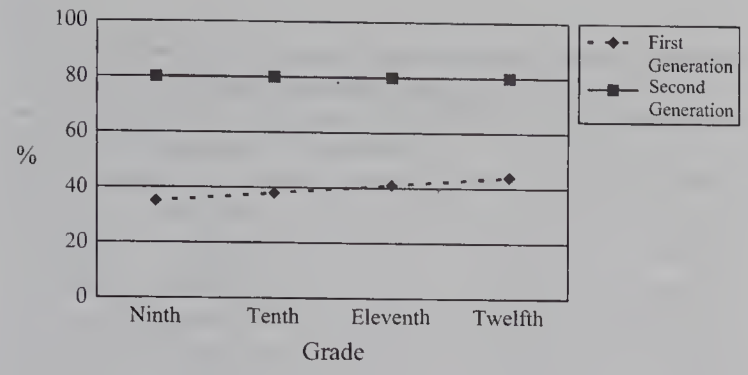


Figure 1. Change over time in likelihood of including a pan-ethnic or an American term in adolescents'ethnic label according to generation.

this analysis, generation in Equations 4 and 5 was replaced with age of immigration, which was centered at the mean of the sample. With older ages of immigration, participants were less likely to choose a pan-ethnic or an American label ($b = -.03, p < .05$). However, fluctuations in ethnic labeling over time did not significantly vary according to age of immigration.

Within-Person Changes in Ethnic Labeling

Amount of within-person change. Although there was not a normative trend regarding the inclusion of a pan-ethnic or an American term in adolescents' ethnic labels over time, there was a substantial amount of change in any direction (i.e., toward or away from inclusion of a pan-ethnic or an American term) within individuals during the 4 years of high school. Overall, 47.8% of the sample changed whether or not they included a pan-ethnic or an American term in their most descriptive label over at least one grade interval during the 4 years of high school. As shown in Figure 2, between 18.2% and 22.3% of the sample changed whether they included a pan-ethnic or an American term in their label at each grade interval. Likelihood of changing whether or not a pan-ethnic or an American term was included in the label varied according to generation across the final grade interval, between 11th and 12th grades, $\chi^2(1, N = 288) = 5.10, p < .05$, such that first-generation adolescents were more likely to change. There were no ethnic differences and no other generation differences in the likelihood of changing whether a pan-ethnic or an American term was included in adolescents' most descriptive label.

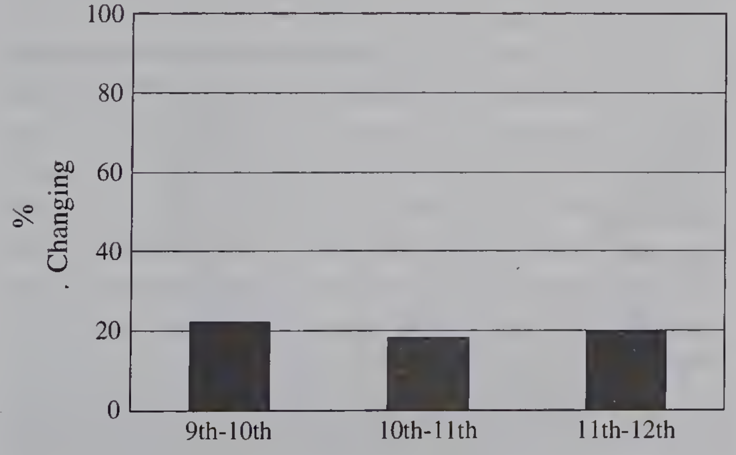


Figure 2. Percentage of adolescents changing as to whether a pan-ethnic or an American term was included in their ethnic label.

Given the degree of changes in ethnic labeling within adolescents across time, an additional set of HLM was estimated to determine whether changes in adolescents' inclusion of a pan-ethnic or an American term from year to year were associated with concurrent changes in their closeness with parents, strength of ethnic identity, and heritage language proficiency. Separate models were estimated for each of the predictor variables (closeness with parents, strength of identity, and heritage language proficiency). Given that there was a maximum of four time points per person and that some participants had only two or three time points, we did not have enough power and degrees of freedom to estimate all the predictors simultaneously. The general form of the model used for these analyses was as follows:

$$\text{Pan-ethnic or American term}_{ij} = b_{0j} + b_{1j}(\text{predictor}) + e_{ij} \tag{6}$$

$$b_{0j} = c_{00} + c_{01}(\text{gender}) + c_{02}(\text{ethnicity}) + c_{03}(\text{generation}) + u_{0j} \tag{7}$$

$$b_{1j} = c_{10} + c_{11}(\text{gender}) + c_{12}(\text{ethnicity}) + c_{13}(\text{generation}) + u_{1j} \tag{8}$$

Equation 6 shows how adolescents' selection of a pan-ethnic or an American term (0 = *no*, 1 = *yes*) on a particular year (i) for a particular individual (j) was modeled as a function of the average amount of inclusion of the term by the individual (b_{0j}) and the specific predictor variable (b_{1j}; i.e., closeness with parents, strength of ethnic identity, or heritage language proficiency). Closeness with parents and strength of identity were coded from 0 to 4 and heritage language proficiency was coded from 0 to 5. Equations 7 and 8 show how the average amount of inclusion of the term and the effect of the predictor variable were modeled as a function of gender, ethnicity, and generational status, which were coded in the same manner as before.

As before, an additional analysis was conducted with first-generation participants, in which generation in Equations 7 and 8 was replaced with age of immigration. In no case, did the relationship between the predictor and the selection of a pan-ethnic or an American label vary according to age of immigration.

Closeness with parents. As shown in Table 3, there was no association between changes in closeness with parents and changes in the inclusion of an American or pan-ethnic term in adolescents' most descriptive ethnic label ($b = .00$, *ns*).

Strength of ethnic identity. Separate analyses were performed using ethnic search and ethnic affirmation. As shown in Tables 4 and 5, lower levels of search and affirmation were each associated with a greater likelihood of adolescents including a pan-ethnic or an American term in their most descriptive ethnic label each year (search: $b = -.04$, $p < .01$; affirmation: $b = -.08$, $p < .001$). However, the relationship between affirmation and label choice was moderated by generation such that the association was significant for first-generation participants ($b = -.13$, $p < .001$) but not for second-generation participants ($b = -.03$, *ns*). The analysis examining the relationship between affirmation and likelihood of including a pan-ethnic or an American label in one's most descriptive label was the one instance in which the results varied between HLM and HGLM. In the HGLM, there was a significant relationship between affirmation and label choice that was not moderated by generation.

Heritage Language Proficiency

As shown in Table 6, lower levels of heritage language proficiency were associated with a greater likelihood of including a pan-ethnic or an American label in one's most descriptive label ($b = -.05$, $p < .01$) each year. However, this relationship was moderated by generation such that the association was significant for first-generation participants ($b = -.08$, $p < .01$) but not for second-generation participants ($b = -.01$, *ns*).

Parental Education

A final set of analyses was conducted using parental education as an additional predictor in Equations 4, 5, 7, and 8 to determine whether it was associated with any of the patterns reported earlier and whether it might explain any observed ethnic or generational differences. Parental education was a significant predictor in only one of the analyses: Among adolescents with more highly educated parents, the negative association between the heritage language proficiency

Table 3

Hierarchical Linear Models Predicting Inclusion of a Pan-Ethnic or an American Term According to Closeness With Parents

	Pan-ethnic or American term, *b* (*SE*)
Intercept	.61 (0.06)***
Gender	−.02 (0.05)
Ethnicity	.04 (0.05)
Generation	.11 (0.06)
Closeness with parents	.00 (0.03)
Gender	.00 (0.02)
Ethnicity	−.01 (0.02)
Generation	.04 (0.03)

Note. Gender was coded as *boys* = −1, *girls* = 1; ethnicity was coded as *Latino* = −1, *Asian* = 1; generation was coded as *first generation* = −1, *second generation* = 1. All predictors were uncentered.
*$p < .05$. **$p < .01$. ***$p < .001$.

Table 4

Hierarchical Linear Models Predicting Inclusion of a Pan-Ethnic or an American Term According to Ethnic Identity Search

	Pan-ethnic or American term, *b* (*SE*)
Intercept	.70 (0.04)***
Gender	.00 (0.03)
Ethnicity	.02 (0.03)
Generation	.16 (0.04)***
Search	−.04 (0.01)**
Gender	.00 (0.01)
Ethnicity	.00 (0.01)
Generation	.02 (0.01)

Note. Gender was coded as *boys* = −1, *girls* = 1; ethnicity was coded as *Latino* = −1, *Asian* = 1; generation was coded as *first generation* = −1, *second generation* = 1. All predictors were uncentered.
**$p < .01$. ***$p < .001$.

Table 5

Hierarchical Linear Models Predicting Inclusion of a Pan-Ethnic or an American Term According to Affirmation and Belonging

	Pan-ethnic or American term, *b* (*SE*)
Intercept	.84 (0.05)***
Gender	−.05 (0.04)
Ethnicity	−.03 (0.04)
Generation	.06 (0.05)
Affirmation	−.08 (0.02)***
Gender	.02 (0.01)
Ethnicity	.01 (0.01)
Generation	.05 (0.02)**

Note. Gender was coded as *boys* = −1, *girls* = 1; ethnicity was coded as *Latino* = −1, *Asian* = 1; generation was coded as *first generation* = −1, *second generation* = 1. All predictors were uncentered.
*$p < .05$. **$p < .01$. ***$p < .001$.

Table 6
Hierarchical Linear Models Predicting Inclusion of a Pan-Ethnic or an American Term According to Heritage Language Proficiency

	Pan-ethnic or American term, *b* (*SE*)
Intercept	.77 (0.06)***
Gender	−.07 (0.04)
Ethnicity	.01 (0.04)
Generation	.07 (0.06)
Heritage language proficiency	−.04 (0.02)**
Gender	.02 (0.01)
Ethnicity	−.01 (0.01)
Generation	.04 (0.02)*

Note. Gender was as coded *boys* = −1, *girls* = 1; ethnicity was coded as *Latino* = −1, *Asian* = 1; generation was coded as *first generation* = −1, *second generation* = 1. All predictors were uncentered.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

and the inclusion of an American term was reduced ($b = .02, p < .05$). In no case, however, did the inclusion of parental education change any of the results regarding ethnicity or generation reported earlier.

Discussion

Adolescents from immigrant families showed no normative trend toward the inclusion of pan-ethnic and American terms in their most important ethnic labels across the high school years. Generational changes remained stable during high school, with second-generation students being more likely to use pan-ethnic and American terms than first-generation students, testifying to the continued importance of birthplace in adolescents' choice of the ethnic labels with which they identify most closely (Rumbaut, 1994). Even after spending more years in American society, many foreign-born adolescents likely feel unwilling or unable to identify primarily with the ethnic labels common to the United States because they were born in other countries, such as Mexico or China. Second-generation adolescents, in turn, may feel unable to consider themselves primarily "Mexican" or "Chinese" because they were not born in those countries. Although first-generation adolescents who arrived in the United States at earlier ages were more likely to use pan-ethnic and American terms on average, the lack of change over time according to age of immigration suggests that it is the age at which the adolescents enter the United States that makes a difference rather than the number of years they have spent in the country. Ethnic differences also remained stable across time. The trend for adolescents from Asian immigrant families to more likely use an American term in their most important ethnic label may be due to a greater identification with American society resulting from their higher level of academic success and their families' relatively easier economic integration as compared to those from Latin American backgrounds (Portes & Zhou, 1993).

The lack of normative change toward the inclusion of pan-ethnic and American terms in adolescents' most important label demonstrates how differences across different generations do not necessarily mean that more time in the United States will produce changes within immigrants themselves (Fuligni, 2001). At least during the high school years, adolescents from immigrant families as a whole do not show changes that would imply acculturation or assimilation to the dominant ethnic categories and labels used in American society. It is possible that the stability may be a function of the relative constancy of the social environments of the students during high school. The contexts of the adolescents' lives show little change between the 9th and 12th grades as they attend the same schools with generally the same peers and teachers. Instead, normative changes in ethnic labeling may be more likely to occur at periods of social transition, such as the move out of high school and into young adulthood. As young adults gain new experiences through work or education, they may become more deeply aware of their ethnic background (Phinney, 2003). Practically speaking, young adults often use their newly found independence and their college and work experiences to immerse themselves in diverse perspectives, student or professional organizations, and coursework to explore and learn more about their identity (Tatum, 1997). It would be important in future work to follow adolescents from immigrant families during such a significant period of developmental transition, although it is unlikely that generational differences would ever disappear completely as studies of adults have found consistent generational differences in ethnic labeling (Lien et al., 2003; Masuoka, 2006).

Although there was great stability in the overall numbers of adolescents who chose different ethnic labels during high school, there were changes within adolescents themselves from year to year. Therefore, rather than a period of great stability, the high school years are a time of fluctuation in the ethnic labels with which adolescents most closely identify. The fluctuation is not in any one direction. Rather, adolescents from immigrant families appear to be trying on different ethnic labels each year, and similar numbers of students move toward and away from the use of a pan-ethnic or an American term. Such fluctuation would be consistent with the portrayal of the adolescent years as a key period in the developmental of ethnic identity, which usually involves a process of search and exploration of one's cultural background and place in American society (Phinney, 1990).

Rather than being random fluctuation, however, the changes in most important ethnic labels within adolescents themselves were significantly associated with changes in other significant aspects of the adolescents' psychological and social development thought to be central to the larger acculturation process of adolescents from immigrant families (Phinney, 2003; Portes & Rumbaut, 2002, Tseng & Fuligni, 2001). Changes in the strength of adolescents' ethnic identity predicted changes in their ethnic labeling, with lower levels of ethnic search and affirmation being associated with the greater probability of including pan-ethnic and American terms. The process of ethnic identity development, although it takes place in a social context in which pan-ethnic and American terms are dominant, likely requires adolescents from immigrant families to consider their national heritage. Retaining a national origin label such as "Mexican" or "Chinese" may require more identity work and effort, which would be reflected in higher levels of ethnic search and affirmation. The inclusion of pan-ethnic and American terms, in contrast, could partially be a result of simply accepting the ethnic categories that American society places adolescents into through school forms and other official documents that allow for identification with only pan-ethnic or hyphenated-American labels. Accepting these labels, therefore, may require lower levels of ethnic search and affirmation.

Change in adolescents' heritage language proficiency was associated with changes in their most important ethnic labels such that lower levels of proficiency were associated with a greater probability of including pan-ethnic and American terms. Language use and proficiency have been shown to be linked to the strength of adolescents' ethnic identity in previous research (Phinney et al., 2001). The ability to speak their families' heritage language, such as Spanish or Chinese, likely enables adolescents to explore their families' cultural and national heritage more deeply through communication with family members and consumption of non-English information sources and media. Adolescents from immigrant families also may feel that they cannot consider themselves primarily "Mexican" or "Chinese" if they cannot speak the language very well. Continued research should more thoroughly explore the meaning of language use and ability for ethnic labeling and identity through the use of qualitative methods such as in-depth personal interviews and focus groups.

Interestingly, the link between changes in heritage language ability and changes in the inclusion of a pan-ethnic or an American label varied such that it was stronger for first-generation adolescents as compared to second-generation adolescents. This may be because of the high level of heritage language ability of the first-generation adolescents, who were born outside of the United States. As a result, heritage language ability may have more meaning for these teenagers, and it may be seen as a greater indicator of identification with and connectedness to the family's country and culture of origin. For second-generation adolescents, heritage language ability appears to be less of a meaningful indicator for ethnic and cultural identity.

In contrast to the strength of ethnic identity and heritage language proficiency, changes in closeness with parents were not predictive of changes to ethnic labeling. It may be that ethnic labeling is more dependent upon social and psychological factors most closely tied to ethnicity and culture than more generic aspects family relationships. In addition, it is important to note that the associations of strength of ethnic identity and heritage language proficiency with ethnic labeling were observed *within* adolescents themselves, thereby ruling out potential confounding factors associated with traditional correlations observed between adolescents. Rather, the associations observed in this study suggest that these social processes are closely linked to one another among adolescents from immigrant families.

The results of this study highlight how adolescents from immigrant families appear to be exploring and experimenting with different ethnic labels across the high school years, and it is important in continuing work to examine how these explorations vary across different social contexts and settings. The ethnic context of adolescents' neighborhoods and schools should make a difference in the labels that they use to describe themselves. Adolescents who participated in this study all lived in contexts in which there are substantial numbers of adolescents from similar immigrant and ethnic backgrounds, thereby potentially leading toward lower rates of inclusion of pan-ethnic and American terms in their ethnic labels. The present study was unable to determine whether variation in the ethnic and immigrant concentration across schools and neighborhoods would make a difference in adolescents' ethnic labels, but it is possible that higher levels of pan-ethnic and American labeling would be observed among immigrants in contexts with few other coethnics. Given the increased dispersal of immigrants to nontraditional receiving locations in the United States, this would be an important topic for future research.

Other studies have suggested that the use of different ethnic labels may vary contemporaneously across different proximal social settings, such as school, home, and peer groups (Harris & Sim, 2002; Moje & Martinez, 2007). It would be useful in future research to combine the approach taken in this study with methods that assess multiple settings to obtain a fuller understanding of the explorations and fluctuations that take place in ethnic labeling during the adolescent years.

Additional studies also can examine the role of other psychological and social factors in adolescents' choice of ethnic labels. Experience with discrimination has been suggested to play a role in the ethnic identification of adolescents from immigrant families, yet the direction of the effect is unclear. Waters (1999) reported that adolescents who did not identify with their parents' West Indian cultural background and instead identified with labels such as Black and African American tended to report more frequent experiences with discrimination. In contrast, Portes and Rumbaut (2001) found that adolescents from immigrant families who identified with their parents' national origin reported the highest levels of discrimination. A clearer understanding of the role of discrimination could be obtained by better analyzing the precise social meanings and social positioning of individual ethnic labels (e.g., Mexican vs. Latino vs. Chinese vs. Asian), as well as the ways in which adolescents from immigrant families themselves make sense of those social meanings. In addition to examining discrimination, closer analyses of adolescents' friendships, attitudes toward American society, and involvement in cultural activities and organizations would be important to examine.

In summary, although ethnic labeling showed no normative trend of acculturation or assimilation, the high school years represent a period of fluctuation in the labels that adolescents from immigrant families identify with most closely. These fluctuations appear to be tied to fluctuations in other significant aspects of their psychological and social development, and future research should expand the developmental periods, social settings, and potential predictors being assessed to better understand a process that is an important component of the acculturation and development of adolescents from immigrant families.

References

Ashmore, R. D., Deaux, K., & McLaughlin-Volpe, T. (2004). An organizing framework for collective identity: Articulation and significance of multidimensionality. *Psychological Bulletin, 130*, 80–114.

Bowman, P. J., & Howard, C. (1985). Race-related socialization, motivation, and academic achievement: A study of Black youths in three-generation families. *Journal of the American Academy of Child Psychiatry, 24*, 134–141.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.

Buriel, R., & Cardoza, D. (1993). Mexican American ethnic labeling: An intrafamilial and intergenerational analysis. In M. Bernal & G. Knight (Eds.), *Ethnic identity* (pp. 197–210). Albany: State University of New York Press.

California Department of Education. (2006). *School summary data*. Retrieved January 27, 2006, from http://dq.cde.ca.gov/dataquest/

Doan, G. O., & Stephan, C. W. (2006). The functions of ethnic identity: A New Mexico Hispanic sample. *International Journal of Intercultural Relations, 30*, 229–241.

Fuligni, A. J. (1998). Parental authority, adolescent autonomy, and parent-adolescent relationships: A study of adolescents from Mexican, Chinese, Filipino, and European backgrounds. *Developmental Psychology, 34*, 782–792.

Fuligni, A. J. (2001). A comparative longitudinal approach to acculturation among children from immigrant families. *Harvard Educational Review, 71*, 566–578.

Fuligni, A. J., Witkow, M., & Garcia, C. (2005). Ethnic identity and the academic adjustment of adolescents from Mexican, Chinese, and European backgrounds. *Developmental Psychology, 41*, 799–811.

Gonzales, N. A., Deardorff, J., Formoso, D., Barr, A., & Barrera, M. J. (2006). Family mediators of the relation between acculturation and adolescent mental health. *Family Relations, 55*, 318–330.

Harris, D. R., & Sim, J. J. (2002). Who is multiracial? Assessing the complexity of lived race. *American Sociological Review, 67*, 614–627.

Lien, P., Conway, M. M., & Wong, J. (2003). The contours and sources of ethnic identity choices among Asian Americans. *Social Science Quarterly, 84*, 461–481.

Masuoka, N. (2006). Together they become one: Examining the predictors of panethnic group consciousness among Asian Americans and Latinos. *Social Science Quarterly, 87*, 993–1011.

Matute-Bianchi, M. E. (1991). Situational ethnicity and patterns of school performance among immigrant and non-immigrant Mexican-descent students. In M. A. Gibson & J. U. Ogbu (Eds.), *Minority status and schooling: A comparative study of immigrant and involuntary minorities* (pp. 205–248). New York: Garland.

Moje, E. B., & Martinez, M. (2007). The role of peers, families, and ethnic-identity enactments in educational persistence and achievement of Latino and Latina youths. In A. J. Fuligni (Ed.), *Contesting stereotypes and creating identities: Social categories, social identities, and educational participation* (pp. 209–238). New York: Russell Sage Foundation.

Olson, D. H., Sprenkle, D. H., & Russell, C. S. (1979). Circumplex model of marital and family systems: I. Cohesion and adaptability dimensions, family types, and clinical applications. *Family Process, 18*, 3–28.

Phinney, J. S. (1990). Ethnic identity in adolescents and adults: A review of research. *Psychological Bulletin, 108*, 499–514.

Phinney, J. S. (1992). The Multigroup Ethnic Identity Measure: A new scale for use with diverse groups. *Journal of Adolescent Research, 7*, 156–176.

Phinney, J. S. (2003). Ethnic identity and acculturation. In K. M. Chun, P. B. Organista, & G. Marin (Eds.), *Acculturation: Advances in theory, measurement, and applied research* (pp. 63–81). Washington, DC: American Psychological Association.

Phinney, J. S., Romero, I., Nava, M., & Huang, D. (2001). The role of language, parents and peers in ethnic identity among adolescents in immigrant families. *Journal of Youth and Adolescence, 30*, 135–153.

Portes, A., & Rumbaut, R. G. (2001). *Legacies: The story of the immigrant second generation*. New York: Russell Sage Foundation.

Portes, A., & Zhou, M. (1993). The new second generation: Segmented assimilation and its variants. *Annals of the American Academy of Political and Social Sciences, 530*, 74–96.

Ruble, D. N., Alvarez, J., Bachman, M., Cameron, J., Fuligni, A. J., García Coll, C., et al., (2004). The development of a sense of "we": The emergence and implications of children's collective identity. In M. Bennett & F. Sani (Eds.), *The development of the social self*. East Sussex, England: Psychology Press.

Rumbaut, R. G. (1994). The crucible within: Ethnic identity, self-esteem and segmented assimilation among children of immigrants. *International Migration Review, 28*, 748–794.

Tatum, B. D. (1997). *Why are all the Black kids sitting together in the cafeteria? And other conversations about race*. New York: Basic Books.

Tseng, V., & Fuligni, A. J. (2000). Parent-adolescent language use and relationships among immigrant families. *Journal of Marriage and the Family, 62*, 465–476.

Waters, M. C. (1999). *Black identities: West Indian immigrant dreams and American realities*. New York: Russell Sage Foundation.

Zhou, M. (1997). Growing up American: The challenge confronting immigrant children and children of immigrants. *Annual Review of Sociology, 23*, 63–95.

Appendix

Ethnic Labels Presented to Adolescents for Selection

African-American
American
Asian
Asian-American
Black
Brazilian
Cambodian
Central-American
Chicano/a
Chinese
Chinese-American
European-American
El Salvadoran
Filipino/a
Guatemalan
Hispanic
Hispanic-American
Indian
Iranian
Japanese
Japanese-American
Korean
Korean-American
Latino/a
Latino-American
Mexican
Mexican-American
Middle Eastern
Native American
Nicaraguan
Nicaraguan-American
Pacific Islander
Pakistani
Persian
Salvadoran-American
Samoan
Spanish
Spanish-American
Thai
Taiwanese
Taiwanese-American
Vietnamese
Vietnamese-American
White
Australian[a]
Asian Indian[a]
Cantonese[a]
Burmese[a]
Peruvian[a]
Uruguayan[a]
Indonesian[a]
Puerto Rican[a]

Note. Labels were presented to adolescents in alphabetical order, without their categorization.

[a] Labels that were not given to students as options but were written in by the students from Asian and Latin American immigrant families as labels that describe them.

Child Development, July/August 2008, Volume 79, Number 4, Pages 957–978

A Developmental Perspective on Full- Versus Part-Day Kindergarten and Children's Academic Trajectories Through Fifth Grade

Elizabeth Votruba-Drzal
University of Pittsburgh

Christine P. Li-Grining
Loyola University Chicago

Carolina Maldonado-Carreño
University of Pittsburgh

Children's kindergarten experiences are increasingly taking place in full- versus part-day programs, yet important questions remain about whether there are significant and meaningful benefits to full-day kindergarten. Using the Early Childhood Longitudinal Study's Kindergarten Cohort (N = 13,776), this study takes a developmental approach to examining associations between kindergarten program type and academic trajectories from kindergarten (ages 4–6 years) through 5th grade (ages 9–12 years). Full-day kindergarten was associated with greater growth of reading and math skills from fall until spring of kindergarten. Initial academic benefits diminished soon after kindergarten. The fade-out of the full-day advantage is in part explained by differences in the children who attend part- and full-day kindergarten as well as school characteristics.

Over the past 35 years, kindergarten programs in the United States have been transformed. When states and localities began implementing publicly funded kindergarten in the 1960s and 1970s, they consisted mostly of part-day programs and focused on easing young children's transitions to school by providing opportunities to socialize and play in group settings. Fewer than 15% of all 5-year-olds attended full-day programs in the 1970s (Elicker & Mathur, 1997). In recent years, the number of children attending full-day kindergarten programs has nearly doubled. At the beginning of the 1980s, just over 25% of U.S. children in kindergarten attended full-day programs (Burriss, 2000). By the end of the 1990s, this number grew to more than 55% (Walston & West, 2004).

As the number of children in full-day kindergarten programs has grown, so too has research comparing the effects of full- and part-day kindergarten on children's academic achievement (for a review, see Lee, Burkam, Ready, Honigman, & Meisels, 2006). Existing research on the benefits of full-day kindergarten has largely been conducted outside the developmental literature, where studies have focused mostly on the educational aspects of full-day kindergarten (Lee et al., 2006) or the policy debates surrounding full-day kindergarten (Cannon, Jacknowitz, & Painter, 2006). Though effects on children's development are of primary interest in studies of full-day kindergarten programs, developmental models and analytic techniques have rarely been used in research comparing full- and part-day kindergarten. Grounded in bioecological theory, the goal of this study was to contribute a developmental perspective to the literature on kindergarten program type (Bronfenbrenner, 1979; Bronfenbrenner & Morris 2006). Using data from the Early Childhood Longitudinal Study's Kindergarten Cohort (ECLS–K), we examine whether kindergarten program type is linked to initial gains in children's achievement trajectories, after taking into account important characteristics of out-of-school contexts and heterogeneity in the development of academic skills. Furthermore, we consider how long into the elementary school years the advantage of full- versus part-day kindergarten lasts.

Are There Benefits to Attending Full-Day Kindergarten?

Though some investigations on the benefits of attending full-day kindergarten have not found

---

The research reported in this article was made possible by a fellowship from the Spencer Foundation. The views expressed are our own and do not necessarily reflect the views of the Spencer Foundation. We thank Emma Adam, Lindsay Chase-Lansdale, Tom Cook, Greg Duncan, Bob Pianta, Fred Morrison, and Sean Reardon for their comments on prior drafts of this manuscript. We also thank Mathilda du Toit for technical support she provided for the HLM software. Any errors that remain are ours. A special thank you is also extended to the children and families who participated in the Early Childhood Longitudinal Study's Kindergarten Cohort.

Correspondence concerning this article should be addressed to Elizabeth Votruba-Drzal, Department of Psychology, University of Pittsburgh, 4123 Sennott Square, 210 South Bouquet Street, Pittsburgh, PA 15260. Electronic mail may be sent to evotruba@pitt.edu.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7904-0011

differences between children enrolled in full- versus part-day programs, more studies have documented significant benefits of full-day kindergarten for children's academic skills (for a review, see Lee et al., 2006). When compared to children attending part-day kindergarten, those attending full-day kindergarten tend to perform better on tests of reading, math, and science achievement, and have lower levels of special education placements and grade retention (Cannon et al., 2006; Clark & Kirk, 2000; Cryan, Sheehan, Wiechel, & Bandy-Hedden, 1992; Elicker & Mathur, 1997; Entwisle, Alexander, Cadigan, & Pallas 1987; Gullo, 2000; Gullo, Bersani, Clements, & Bayless, 1986; Kaplan, 2002; Karweit, 1992; Lee et al., 2006; Walston & West, 2004; Weiss & Offenberg, 2002). Overall, these advantages tend to be small to moderate.

A Developmental Approach to Studying Full-Day Kindergarten

Noticeably absent from existing research on the benefits of full-day kindergarten have been the voices of developmental scholars, which is rather surprising, given the developmental nature of this policy question. At the core of discussions about kindergarten program type is the question of whether full-day kindergarten is associated with meaningful differences in children's individual achievement trajectories during the kindergarten year and beyond, net of other important contexts that shape their development. Developmental theory and analytic techniques are poised to make an important contribution to this policy discussion (Foster & Kalil, 2005). Developmental science highlights both the multicontextual nature of environmental influences on children's lives and the heterogeneity of developmental trajectories. By incorporating these two core tenets of developmental psychology into our analysis of kindergarten program type, we hope to strengthen our understanding of the benefits of full-day kindergarten attendance to children's academic achievement.

Recognizing multiple contexts of children's development. Almost 30 years ago, Bronfenbrenner (1979) advanced the field of child development by introducing Ecological Systems Theory, which recognizes the embeddedness of children's lives in multiple contexts. In what subsequently became known as the bioecological model of human development, Bronfenbrenner argued for the study of child development in ecological context or, as he stated so clearly, "in the actual environments in which human beings lived their lives" (Bronfenbrenner & Morris, 2006, p. 794). Although children's experiences in school play a prominent role in shaping their development throughout middle childhood, other contexts including families and nonparental care settings have important influences on children's development. Yet, studies comparing full- versus part-day kindergarten have not sufficiently taken into account the influences of these out-of-school settings.

Studies have repeatedly demonstrated that families represent the most influential context in children's lives (Coleman, 1966; National Institute of Child Health and Human Development [NICHD] Early Child Care Research Network, 2002). An important aspect of families' lives that has received notably less attention in the literature on full-day kindergarten is the cognitive stimulation parents provide for their children. Yet, learning experiences in the context of the home environment play a central role in shaping the development of children's academic skills during the preschool and early elementary school years (Morrison, Bachman, & Connor, 2005). Global measures of cognitive stimulation in the home environment have been linked to children's vocabulary, reading, and math skills (Bradley, Corwyn, Burchinal, McAdoo, & Garcia Coll, 2001; Smith, Brooks-Gunn, & Klebanov, 1997). These measures cover a broad range of learning experiences including the frequency with which parents read to children; teach them letters, numbers, and shapes; and provide them with cognitively stimulating toys.

One salient dimension of families' lives that has been addressed in research on full-day kindergarten is the economic condition of children's households (e.g., Cannon et al., 2006). One fifth of young children in the United States live in poverty (National Center for Children in Poverty, 2006), which prior studies have consistently identified as a risk factor for children's academic and cognitive skills (Dearing, McCartney, & Taylor, 2001; Duncan & Brooks-Gunn, 1997; McLoyd, 1998; Patterson, Kupersmidt, & Vaden, 1990; Shonkoff & Phillips, 2000; Votruba-Drzal, 2006). Children from economically disadvantaged households tend to experience less consistent, supportive, and cognitively stimulating caregiving than those from middle- and upper-class families (Votruba-Drzal, 2003, 2006). Moreover, children from low-income households face a variety of risk factors that threaten their health and development. They are, for example, more likely to be raised in a single-parent household, have a teenage mother, experience family instability, and grow up in a neighborhood characterized by high levels of violence (Brooks-Gunn, Klebanov, & Liaw, 1995; Evans & English, 2002; Rutter, 1981). Particularly detrimental to children's academic competence is economic hardship that occurs during early childhood, as economic gaps in early academic achievement tend to continue and to exacerbate in middle childhood (Alexander, Entwisle, & Dauber, 1993; Duncan & Brooks-Gunn, 1997; Duncan, Brooks-Gunn, Yeung, & Smith, 1998; Entwisle, Alexander, & Olson, 1997; Votruba-Drzal, 2006).

Beyond the home environment, child-care settings have become increasingly important contexts for children's development, as women have entered the labor force in higher numbers. Several characteristics of child-care settings have been linked to the development of children's academic skills over time, including the number of hours per week children are in care (Bogenschneider & Steinberg, 1994; Broberg, Wessels, Lamb, & Hwang, 1997; Brooks-Gunn, Han, & Waldfogel, 2002; Field, 1991; Harvey, 1999; NICHD, 2000; Votruba-Drzal, Coley, & Chase-Lansdale, 2004) and the type of care that children experience, with center-based care being linked to better academic performance than home-based settings and parental care (Broberg et al., 1997; Caughy, DiPietro, & Strobino, 1994).

Despite the roles home and child-care environments play in shaping children's development, these out-of-school contexts have been left largely unexamined in studies of full- versus part-day kindergarten. This may be especially problematic in regard to studying kindergarten program type because part-day kindergarteners likely spend the majority of their time in home and child-care settings. If children's experiences outside school are related to their academic skills and their parents' selection of kindergarten, then the failure to take these important contexts into account when considering the effects of full- versus part-day kindergarten may produce misleading results.

Mapping trajectories of achievement. A central concept in developmental science is the notion that children's development is characterized by unique trajectories, with some children developing at faster rates than others across time (National Research Council and Institute of Medicine, 2000). A handful of existing studies that have followed children beyond kindergarten suggest that although there are initial benefits to full-day kindergarten, these advantages do not persist after the early years of elementary school. These studies show that the academic disadvantages of part-day kindergarten fade out sometime between first and third grade (Karweit, 1987; Ohio State Legislative Office of Education, 1997; Weiss & Offenberg, 2002). For example, in a recent study using nationally representative data and rigorous analytic techniques, Cannon et al. (2006) found that full-day kindergarten participation was related to modestly higher reading and math scores at the end of the kindergarten year when compared to part-day kindergarten enrollment. After the kindergarten year, however, part-day kindergarten children caught up to their full-day kindergarten counterparts. The disadvantage of attending part-day kindergarten was greatly diminished by first grade and completely eliminated by the spring of third grade.

Much of the literature examining associations between kindergarten program type and academic achievement has been limited to two repeated assessments and has, therefore, not been able to account for the variability in children's growth in academic achievement when comparing the achievement outcomes of part-day kindergarteners to full-day kindergarteners. Furthermore, a focus on two repeated assessments cannot adequately capture the shape of children's academic trajectories. In short, research in this area has conceived of child development as occurring uniformly and incrementally across children. Indeed, nearly all studies of kindergarten program type have relied on the analysis of two waves of achievement scores (e.g., change models or residualized change models) taken before and after exposure to full- versus part-day kindergarten to identify associations between children's full- versus part-day kindergarten attendance and academic achievement. In doing so, studies have conceptualized full- versus part-day kindergarten's effects on children's development as providing incremental gains to achievement that are accrued between the beginning and the end of the kindergarten year and have not been able to provide detailed information about when the advantages of full-day kindergarten attendance subside.

Though the analysis of change scores has benefits for reducing potential sources of bias (Allison, 1990; Willett, 1988), they are characterized by significant limitations that have been reviewed extensively in the literature (Rogosa, Brandt, & Zimowski, 1982; Singer & Willett, 2003; Willett, 1988). Among the most significant limitations are that these approaches do not recognize two central concepts in developmental science: that child development is heterogeneous and that patterns of child development are more accurately portrayed by careful estimation of trajectories of individual growth rather than by simple change scores.

Over the past 20 years, there has been a reorientation in developmental research from what Willett (1988) refers to as a focus on *change* to more nuanced models of *growth*. Advances such as hierarchical linear modeling (HLM) have paved the way for more precise estimation of children's individual-specific growth trajectories (Burchinal & Appelbaum, 1991; Burchinal, Nelson, & Poe, 2006; Raudenbush & Bryk, 2002; Singer & Willett, 2003). These approaches to measuring development, however, have not yet been applied to the question of whether full-day kindergarten has meaningful implications for children's achievement trajectories. Nevertheless, these methods are an integral part of a developmental perspective because they recognize the heterogeneous nature of development. Here, we specifically use HLM to determine whether full-day kindergarten enrollment is related to steeper growth in individual-specific achievement trajectories during the kindergarten year and to gain a more precise estimation of the duration of the full-day kindergarten advantage.

Research Goals

Individual differences in children's achievement trajectories, across the kindergarten year and beyond, stem from a unique constellation of child, school, home, and child-care characteristics (National Research Council and Institute of Medicine, 2000; Pianta, Rimm-Kaufman, & Cox, 1999). Guided by a developmental approach that highlights this heterogeneity and contextual embeddedness of children's lives, the aim of this study was to examine whether kindergarten program type explains individual differences in children's academic trajectories from kindergarten through fifth grade, net of salient aspects of children's home and child-care experiences. This study begins by providing a rich description and comparison of children attending full- and part-day kindergarten. Next, it considers whether full-day, as opposed to part-day, kindergarten attendance is linked with greater initial growth of reading and math skills during the kindergarten year after taking into account the influences of important characteristics of children's home and child-care settings. It then considers how long these benefits are sustained by examining trajectories of student achievement from the spring of kindergarten through the end of fifth grade.

Method

The data for this study come from the Early Childhood Longitudinal Study–Kindergarten Class of 1998–1999 (ECLS–K), a nationally representative cohort of children (U.S. Department of Education, National Center for Education Statistics, 2001). The ECLS–K was designed to study relations among children's developmental trajectories and their family, preschool, and school experiences. The sample was selected to be nationally representative of all children entering kindergarten in the fall of 1998 using a multistage probability sample design, where the primary sampling units (PSU) were geographic areas consisting of counties or groups of counties, the second-stage sampling units were schools within PSUs, and the final stage sampling units were students within schools. In total, nearly 22,000 kindergarteners throughout the United States participated in the study in the fall of 1998. The ECLS–K collected base-year data in the fall of kindergarten in 1998 and the spring of kindergarten in 1999. Four waves of data have been collected beyond the kindergarten year. These have taken place in the fall and spring of first grade, the spring of third grade, and the spring of fifth grade. The ECLS–K planned to collect data from all children at each wave of the study, with the exception of the fall of first grade, when data were collected only from a representative 30% of the entire sample of schools. Each school year of the survey information was collected from parents, teachers, and school administrators. Parent interviews were conducted by telephone or in person for families without a telephone, and teachers and school administrators were surveyed through self-administered questionnaires. The data were collected across several domains and include multimethod, multisource, in-school assessments of children's cognitive development as well as measures of family, school, and classroom characteristics that have been associated with children's development.

Sample

The sample that was used in this study consists of first-time kindergarteners who remained in the same type of kindergarten program (full or part day) throughout the kindergarten year, who were in kindergarten for at least 4 days per week, and who had at least one valid observation in reading and in math from kindergarten through fifth grade. Only children who had a valid sampling weight (C1PW0) were included in the sample. Altogether, 13,776 children were included in our sample. This represents 78% of the entire sample of children for whom ECLS–K recalibrated the item response theory score (IRT) assessment scores from kindergarten through fifth grade. The number of children dropped from our sample because of our exclusion criteria are as follows: children who repeated kindergarten ($n = 675$), those who switched program type ($n = 240$), those who were not in kindergarten for at least 4 days per week ($n = 389$), those who did not have a valid math or reading assessment ($n = 79$), and those with a sampling weight of 0 or missing weight ($n = 2{,}501$).

Of the 13,776 children included in the sample, 64% and 68% had valid reading and math scores, respectively, for five or six waves of data from the fall of kindergarten through the spring of fifth grade, and 31% and 29% had valid reading and math scores, respectively, for three or four waves of data. An extensive set of 17 covariates are included in our analyses that reflect important characteristics of children and their in-school and out-of-school contexts. Of the 13,776 children in our sample, 87% had valid data on all the 17 covariates. Statistical comparisons of children who had missing child assessments or important covariates showed that they were more disadvantaged across several dimensions. For example, they were more likely to be of an ethnic minority background, scored slightly lower on academic skills, and their parents tended to be less educated and to have incomes at or below the poverty line.

Traditional approaches to handling missing data, such as listwise deletion or mean imputation, have been criticized for biasing estimates, misrepresenting statistical power, and leading to invalid conclusions (Acock, 2005; Rubin, 1987; Widaman, 2006). Therefore, missing data were imputed for the present study using multiple imputation by chained equations (ICE), which was implemented in Stata (Royston, 2004). Multiple imputation was performed in Stata to create five complete data sets that included both the independent and the dependent variables in our models. The primary analysis for this study was performed using HLM 6.04 (Raudenbush, Bryk, Cheong, Congdon, & du Toit, 2004). Following imputation, the five data sets were imported into HLM, which was then used to conduct separate analysis for each set of plausible values. Based on the relative efficiency calculation by Rubin (1987), five multiple imputations were deemed sufficient. Parameter estimates were averaged and standard errors were computed in HLM 6.04 using standard techniques described by Raudenbush et al.

Measures

Academic achievement. Academic achievement in math and reading was measured in school using individualized, direct cognitive assessments designed by the ECLS–K. Several items on the assessment were adapted from existing instruments that have been shown to be reliable and valid measures of children's cognitive and academic development, such as the Woodcock–Johnson Psycho-Educational Battery, the Peabody Picture Vocabulary Test, and the Peabody Individual Achievement Test (Dunn & Dunn, 1981; Dunn & Markwardt, 1970; Markwardt, 1989; Woodcock & Mather, 1989, 1990). Both direct cognitive batteries have been shown to have good internal consistency. The reading assessment tests a broad range of children's language and literacy skills, including letter recognition, receptive vocabulary, and reading comprehension. Items on the mathematics assessment are aimed at measuring general mathematical skills, with questions on topics such as number sense, properties, measurement, and spatial sense.

An important prerequisite for growth modeling is to have an outcome variable measured on a consistent metric over time (Singer & Willett, 2003). To facilitate longitudinal analyses of children's academic achievement, the ECLS–K calculated IRT scores, which estimate children's performance as if they had been administered the whole set of assessment questions. The first set of IRT scores were created for children participating in the kindergarten and first-grade rounds of data collection. As children aged and the assessments were expanded in third and fifth grades, the ECLS–K recalibrated the IRT scores from all prior waves of data to make longitudinal comparisons possible (U.S. Department of Education, National Center for Education Statistics, 2005). Here, we used the recalibrated fifth-grade IRT scores, which is in keeping with prior studies of academic growth using the ECLS–K (e.g., Kaplan, 2002; McCoach, O'Connell, Reis, & Levitt, 2006; Ready, LoGerfo, Burkam, & Lee, 2005).

Kindergarten and school experiences. The primary independent variable of interest in these analyses is kindergarten program type, which is represented with a dichotomous indicator for whether children were enrolled in part- versus full-day kindergarten. Reports of children's kindergarten program type were obtained during interviews with kindergarten teachers. A teacher-reported indicator of whether a child changed teachers during the kindergarten year was included as well. Several basic school demographic characteristics were also included in these analyses, including school type, region, and urbanicity. School type was coded with a dummy variable that indicated whether children attended private versus public schools. The school's geographic location was modeled with dummy variables indicating whether the child lived in the Northeast (omitted from the model as the comparison group), Midwest, South, or West. Finally, schools' urbanicity was represented with variables indicating whether the schools were located in suburbs/large towns (omitted), central cities, or small towns/rural areas.

Out-of-school contexts. This investigation focused on four dimensions of out-of-school contexts: quality of cognitive stimulation in the home, household poverty, and type and extent of child care. First, parents were asked about several different types of learning activities that took place in children's home environments during a typical week. Items were adapted from reliable measures of children's home learning environments that have been widely used and validated in the literature, such as the Home Observation for Measurement of the Environment (HOME) Inventory (Caldwell & Bradley, 1979). For most items, parents rated the frequency with which children engaged in particular learning activities, such as reading books, singing songs, and listening to music, on a 4-point scale (1 = *not at all* and 4 = *3 to 6 times a week*). Dimensions of children's home learning environments that were measured by these items include the level of academic guidance and support provided to children; the degree to which children engaged in new activities and explored and discussed new ideas; and the amount of language stimulation children experienced. Factor analyses were performed to create a composite measure, which was calculated by taking a mean of 24 items asked in the fall and spring of kindergarten (α = .80). Second, family economic resources were represented with a dummy variable indicating whether or not children lived in households with incomes below the poverty line during kindergarten.

The third and fourth aspects of out-of-school contexts were type and extent of child care, which were measured during kindergarten and during the year prior to kindergarten. Prekindergarten type of care was represented categorically by dummy variables indicating whether children were in parental care (omitted), Head Start, other center care, relative care, nonrelative care, or multiple types of care. Type of care during kindergarten was modeled categorically by dummy variables reflecting whether children were in parental care (omitted), center care, relative care, nonrelative care, or multiple types of care. The average hours per week of care during kindergarten and the year prior to kindergarten were each represented continuously.

Family and child characteristics. Several family and individual characteristics that are important for children's developmental trajectories were included in these analyses to further test the robustness of the linkage between kindergarten program type and children's academic trajectories. All these measures were obtained during parental interviews in kindergarten. Parental education was represented categorically as the highest level of education attained by either parent. Categories include less than a high school diploma (omitted), a high school diploma, a vocational or technical program, bachelor's degree, and advanced degree. Maternal age at first birth was measured in years, and the number of children younger than 18 years living in the household was used as an estimate of parents' household caretaking responsibilities. Parental marital status was represented with dummy variables indicating whether parents were married (omitted), never married, separated/divorced, or widowed. Dummy variables were also included for children who lived with nonbiological or adoptive parents, and for whether a non-English language was the primary language spoken at home.

Finally, children's individual characteristics were obtained during parent interviews in kindergarten. Gender was represented with a dummy variable, and the age at which children started kindergarten was measured in months. Child race/ethnicity was represented with a series of dummy variables reflecting whether children were of a Hispanic, non-Hispanic Black, non-Hispanic White (omitted), non-Hispanic Asian, non-Hispanic Native American, or multiethnic background. An indicator of low birth weight was used to represent whether children weighed less than 5.5 pounds at birth.

Analytic Approach

Setting the stage for this investigation, we first sought to depict a national descriptive portrait of children attending full- versus part-day kindergarten. To test whether there were significant differences between these two groups of children, no constant regression analyses were performed, with the dummy indicator for kindergarten program type as the only independent variable, and achievement measures, family and child characteristics, and school demographic characteristics as dependent variables. When testing for differences in family and child characteristics that were represented categorically (e.g., race, parental education), each category was tested as a separate dependent variable to determine whether there were significant differences by kindergarten program type in the proportion of individuals falling into each particular category. This descriptive analysis was performed in Stata, with Huber–White standard error corrections (Huber, 1967) to take into account the nesting of children within schools, and sampling weights were applied.

Next, to examine whether there were initial and long-term benefits to attending full-day kindergarten, we estimated three-level models (Raudenbush & Bryk, 2002) in HLM 6.04 using full information maximum likelihood estimation and sampling weights

(Raudenbush et al., 2004). At Level 1 were children's achievement scores, which were nested within individual children at Level 2, who were nested within schools at Level 3. Children's academic trajectories were estimated using piecewise growth models. To consider whether full-day kindergarten was initially beneficial to the growth of children's academic skills during the kindergarten year, the first trajectory measured growth from the fall of kindergarten to the spring of kindergarten. To examine how long these initial benefits to achievement growth were sustained, the second trajectory measured growth from the spring of kindergarten to the spring of fifth grade. The form of the Level 1 equation can be seen in Equation 1:

$$Y_{tij} = \pi_{0ij} + \pi_{1ij}\text{Time 1}_{tij} + \pi_{2ij}\text{Time2}_{tij} + \varepsilon_{tij}. \qquad (1)$$

We created Time 1 and Time 2 to track growth across kindergarten and from the spring of kindergarten through the spring of fifth grade, respectively. The time of each assessment was measured as the number of months that had passed since September 1, 1998, because the exact start of the school year was not available in the public release data. There was great variability in the month in which each of the assessments took place; for example, the fall of kindergarten assessments took place between September and December. Therefore, the exact number of months that had passed at the time of the assessment was used to yield greater precision in estimating the growth trajectories. The Time 1 variable was centered at the time of the fall kindergarten assessment, and the Time 2 variable was centered at the time of the assessment at the spring of kindergarten. Thus, academic achievement at time *t* for child *i* in school *j* was modeled as a function of the academic achievement of child *i* in school *j* at the fall of kindergarten (π_{0ij}), his or her per month growth of achievement skills during the kindergarten year (π_{1ij}), and his or her monthly growth rate from the spring of kindergarten until the spring of fifth grade (π_{2ij}).

We began our growth trajectory analyses by estimating unconditional growth models, with the coefficients on the slopes in the Level 1 equation estimated as random. The Level 1 intercept and Level 2 intercept predicting the Level 1 intercept were fixed due to model convergence problems (Dearing et al., 2001). The remaining intercepts of the Level 2 equations were estimated as random, and the slope coefficients of the Level 2 equations were fixed at Level 3.

After finding significant variability in the Level 1 and Level 2 parameters, conditional models were estimated to explain the heterogeneity in trajectories of academic achievement.

Equations 2 through 4 were used at Level 2 to model variability in the Level 1 parameters:

$$\pi_{0i} = \beta_{00j} + \beta_{01j}\text{KT}_{ij} + \beta_{02j}\text{FT}_{ij} + \beta_{03j}C_{ij} + \beta_{04j}F_{ij} + \beta_{05j}\text{CCPK}_{ij} + \beta_{06j}\text{CCK}_{ij} \qquad (2)$$

$$\pi_{1ij} = \beta_{10j} + \beta_{11j}\text{KT}_{ij} + \beta_{12j}\text{FT}_{ij} + \beta_{13j}C_{ij} + \beta_{14j}F_{ij} + \beta_{15j}\text{CCK}_{ij} + r_{1ij} \qquad (3)$$

$$\pi_{2ij} = \beta_{20j} + \beta_{21j}\text{KT}_{ij} + \beta_{22j}\text{FT}_{ij} + \beta_{23j}C_{ij} + \beta_{24j}F_{ij} + r_{2ij}. \qquad (4)$$

Here, the initial level of academic achievement was modeled as a function of kindergarten program type (KT_{ij}), and a series of child (C_{ij}), family (F_{ij}), prekindergarten child care (CCPK_{ij}), and kindergarten child-care (CCK_{ij}) characteristics. The time of the child assessment at the fall of kindergarten (FT_{ij}) was included as a predictor as well because children who were assessed later in the school year tended to have more advanced academic skills than did those who were assessed earlier. All independent variables in the Level 2 equation for the intercept were measured in kindergarten. The two slope parameters were estimated as a function of this same set of independent variables, with the exception that only child-care characteristics in kindergarten were used to predict the kindergarten slope. Also, no child-care characteristics were used to model variability in the coefficient on Time 2 because child-care experiences are less relevant in predicting achievement growth after children are in school full time. Except for the indicator for kindergarten program type (KT_k), which was coded as 1 for part-day programs and 0 for full-day programs, all independent variables at Level 2 were centered on the grand mean for the sample, so that the intercepts at Level 2 represent adjusted means for the average study participant in full-day kindergarten (Raudenbush & Bryk, 2002).

At Level 3, we introduced a set of school-level demographic characteristics. Equations 5 through 7 were used to predict the Level 2 intercepts:

$$\beta_{00j} = \gamma_{000} + \gamma_{001}\text{ST}_j + \gamma_{002}R_j + \gamma_{003}U_j \qquad (5)$$

$$\beta_{10j} = \gamma_{100} + \gamma_{101}\text{ST}_j + \gamma_{102}R_j + \gamma_{103}U_j + u_{10j} \qquad (6)$$

$$\beta_{20j} = \gamma_{200} + \gamma_{201}\text{ST}_j + \gamma_{202}R_j + \gamma_{203}U_j + u_{20j}. \qquad (7)$$

Specifically, school type (ST), geographical region (*R*), and urbanicity (*U*) were used to model the intercepts of the Level 2 equations. However, the Level 2 slope coefficients were fixed at Level 3 and did not include any predictors because we had no a priori hypotheses about how school characteristics would moderate the influence of the Level 2 independent variables.

The conditional growth models were built in three steps. In the first model, academic growth trajectories at Level 1 were estimated as a function of whether children attended part-day kindergarten at Level 2. In the second model, important characteristics of children's out-of-school contexts were taken into account, with quality of cognitive stimulation in the home, household poverty status, and type and extent of child care added as covariates at Level 2. In the third and final model, other family and child characteristics were included at Level 2 and school characteristics added at Level 3 as an additional test of the robustness of the association between kindergarten program type and children's achievement trajectories.

Results

Are There Significant Differences Between Children Attending Full- Versus Part-Day Kindergarten?

Laying the groundwork for our multivariate analyses, we first consider whether significant differences exist between students enrolled in part- and full-day kindergarten. Descriptive information for the sample as a whole as well as for subgroups of children by program type can be found in Table 1. As shown in the top row, 45% ($n = 6{,}202$) of children were enrolled in part-day kindergarten and 55% ($n = 7{,}574$) in full-day kindergarten.

Regarding family and household characteristics, the poverty rate among part-day kindergarten students was 6% lower than that for full-day kindergarten students. Among part-day kindergarten students, the highest level of parental education was slightly greater, the proportion of children from married parent households was higher, and the percentage of children from households with parents who had never been married was lower. Not surprisingly, based on these socioeconomic differences, the quality of cognitive stimulation in the home environments of part-day kindergarten students was about one tenth of 1 *SD* higher than that of their peers. A greater percentage of children in part-day kindergarten (14%) versus full-day kindergarten (9.8%) came from a non-English-speaking household.

Moving on to child care, part- and full-day kindergarteners had different experiences during kindergarten. Children attending part-day kindergarten programs spent more hours in child care, had a higher likelihood of receiving nonrelative home-based child care, and had a lower likelihood of receiving relative home-based child care. In contrast, during the year prior to kindergarten, children in part- and full-day kindergarten had somewhat similar child-care experiences. However, part-day kindergarteners spent about 6 hr less in nonparental care per week, were less likely to have attended Head Start, and more likely to be cared for in a nonrelative home-based child-care setting, when compared to full-day kindergarten students.

In addition to contextual differences, children enrolled in part- versus full-day programs varied in terms of their individual characteristics. The most notable of these differences was in the racial composition of the two groups of children. More specifically, a greater proportion of children in part-day kindergarten were of Hispanic, non-Hispanic White, Asian, and multiracial backgrounds, and a smaller proportion were of non-Hispanic Black and Native American race/ethnicity, in comparison to full-day kindergarteners. These differences were moderate to large in magnitude. For example, 7% of part-day kindergarteners were of non-Hispanic Black ethnicity, whereas 22% of full-day kindergarteners fell into this category. Children in part-day kindergarten entered school when they were slightly younger and less likely to have been of low birth weight compared to children in full-day kindergarten. There were no significant differences in child gender by program type.

Furthermore, children's mean levels of academic achievement at the fall of kindergarten did not differ significantly. By the spring of kindergarten, the math skills of the two groups of children were statistically indistinguishable. Still, children in full-day kindergarten outscored their part-day kindergarten counterparts in reading by one tenth of 1 *SD*. By the fall of first grade, children's achievement did not vary across program type and remained similar, until the spring of third grade, when part-day kindergarteners outscored full-day kindergarteners on the math and reading skills measures by one tenth of 1 *SD*. This disparity grew slightly in the spring of fifth grade. Finally, there were several significant differences in school characteristics by kindergarten program type. For example, a greater proportion of children in full-day kindergarten (54.7%) versus part-day kindergarten (14.1%) attended schools located in the South, whereas a lower proportion of full-day kindergarteners attended schools located in the West, Northeast, and Midwest. Full-day kindergarteners also had somewhat higher rates of private versus

Table 1
Descriptive Statistics by Kindergarten Program Type

Characteristics	Full sample[a] (*N* = 13,776), *M* or % (*SD*)	Part-day kindergarten students (*n* = 6,202), *M* or % (*SE*)	Full-day kindergarten students (*n* = 7,574), *M* or % (*SE*)
Child characteristics			
Male	50.4% (50.0)	51.4% (0.7)	50.4% (0.6)
Race			
Non-Hispanic White***	57.8% (49.4)	63.4% (1.8)	54.0% (1.8)
African American***	14.0% (34.7)	6.9% (0.8)	22.3% (1.6)
Hispanic**	16.9% (37.5)	21.9% (1.6)	15.7% (1.2)
Non-Hispanic Asian*	5.5% (22.7)	3.5% (0.4)	2.6% (0.3)
Native American**	3.0% (17.2)	1.4% (0.2)	3.4% (0.7)
Multiracial**	2.8% (16.5)	3.0% (0.3)	2.1% (0.2)
Low birth weight**	7.0% (25.6)	6.3% (0.4)	7.9% (0.4)
Age of entry***	65.8 (4.1)	65.4 (0.1)	66.1 (0.1)
Changed teacher during kindergarten	5.2% (22.3)	5.0% (0.7)	5.9% (0.6)
Family and household characteristics at kindergarten			
Below poverty level***	19.3% (39.4)	16.8% (1.0)	23.2% (1.1)
Maternal age at first birth (years)**	24.0 (5.5)	24.1 (0.2)	23.3 (0.2)
Highest parental education			
Less than high school	9.0% (28.6)	9.3% (0.8)	10.2% (0.6)
High school degree***	25.4% (43.5)	23.9% (0.9)	29.3% (0.9)
Vocational/technical program	32.6% (46.9)	33.5% (0.9)	32.0% (0.8)
Bachelor's degree**	21.1% (40.8)	21.4% (0.9)	18.2% (0.8)
Advanced degree	11.9% (32.4)	11.9% (0.9)	10.2% (0.8)
Marital status			
Married***	71.1% (45.4)	73.8% (1.0)	66.0% (1.2)
Divorced/separated†	13.0% (33.6)	12.7% (0.6)	14.1% (0.5)
Widowed	0.8% (9.0)	0.7% (0.1)	1.0% (0.1)
Never married***	12.9% (33.5)	11.2% (0.7)	15.7% (0.9)
Adoptive/no biological dad***	2.3% (14.9)	1.6% (0.2)	3.2% (0.3)
Non-English home language**	12.7% (33.3)	14.3% (1.2)	9.8% (0.8)
Number of children in household	2.5 (1.2)	2.5 (0.0)	2.5 (0.0)
Home learning environment**	0.0 (0.4)	0.0 (0.0)	0.0 (0.0)
Child-care characteristics at prekindergarten			
Hours of care per week***	25.4 (21.7)	22.3 (0.4)	28.3 (0.4)
Parental care	18.1% (38.5)	19.0% (0.7)	17.7% (0.7)
Head Start***	9.1% (28.8)	7.4% (0.7)	12.2% (0.9)
Center care	43.8% (49.6)	41.7% (1.2)	43.3% (1.1)
Relative home-based care	13.9% (34.6)	14.5% (0.6)	13.4% (0.5)
Nonrelative home-based care***	10.1% (30.1)	12.4% (0.5)	8.2% (0.5)
Multiple types of care	5.0% (21.9)	5.1% (0.3)	5.2% (0.3)
Child-care characteristics at kindergarten			
Hours of care per week***	9.3 (13.8)	11.4 (0.3)	7.9 (0.2)
Parental care**	51.9% (50.0)	50.0% (0.9)	53.3% (0.8)
Center care	17.3% (37.8)	17.4% (0.9)	16.7% (0.8)
Relative home-based care**	18.7% (39.0)	17.0% (0.7)	20.4% (0.7)
Nonrelative home-based care***	10.0% (30.0)	13.5% (0.6)	7.6% (0.4)
Multiple types of care	2.1% (14.2)	2.1% (0.2)	2.0% (0.2)
Kindergarten characteristics			
Region			
Midwest**	25.6% (43.6)	29.2% (2.4)	19.5% (1.9)
Northeast*	18.6% (38.9)	21.1% (2.2)	15.3% (1.8)
West***	22.7% (41.9)	35.6% (2.6)	10.5% (1.4)

(*Continued*)

Table 1
Continued

Characteristics	Full sample[a] (*N* = 13,776), *M* or % (*SD*)	Part-day kindergarten students (*n* = 6,202), *M* or % (*SE*)	Full-day kindergarten students (*n* = 7,574), *M* or % (*SE*)
South***	33.2% (47.1)	14.1% (1.9)	54.7% (2.5)
Private school**	20.6% (40.4)	10.5% (1.4)	17.0% (1.7)
Urbanicity			
Central city	40.8% (49.2)	40.0% (2.4)	39.1% (2.2)
Rural	20.8% (40.6)	18.7% (2.0)	22.7% (2.1)
Large town	38.5% (48.7)	41.3% (2.4)	38.2% (2.3)
Child outcomes			
Reading			
Fall of kindergarten	29.7 (10.1)	29.3 (0.2)	29.4 (0.3)
Spring of kindergarten**	41.4 (13.8)	40.1 (0.3)	41.5 (0.4)
Fall of first grade	48.6 (17.9)	46.9 (0.7)	48.4 (0.9)
Spring of first grade	72.4 (22.4)	71.6 (0.6)	71.3 (0.6)
Spring of third grade*	118.4 (25.0)	118.8 (0.8)	116.2 (0.9)
Spring of fifth grade**	139.5 (23.2)	140.1 (0.9)	137.3 (1.1)
Math			
Fall of kindergarten	23.1 (8.9)	22.8 (0.2)	22.6 (0.2)
Spring of kindergarten	33.6 (11.6)	32.7 (0.3)	33.4 (0.3)
Fall of first grade	40.6 (13.8)	39.6 (0.7)	40.2 (0.6)
Spring of first grade	58.0 (16.8)	57.6 (0.4)	57.2 (0.5)
Spring of third grade**	92.1 (21.5)	92.8 (0.6)	90.4 (0.6)
Spring fifth grade***	113.5 (21.4)	114.6 (0.8)	111.2 (0.9)

Note. Significance levels reflect the statistical significance of differences between students in part- and full-day kindergarten.
[a] Descriptive statistics for the full-sample are unweighted.
$^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

public school enrollment when compared to part-day kindergartener students. In sum, several differences emerged between children attending full- and part-day kindergarten; however, when differences were detected, they tended to be modest in magnitude.

Does Attending Full-Day Kindergarten Yield Benefits to Children's Academic Achievement?

Moving on to our analyses of children's academic trajectories, we estimated math and reading trajectories from the fall of kindergarten until the spring of fifth grade, in order to address the question of whether full-day kindergarten attendance explained variability in children's development. We began by estimating unconditional growth models, which are given in Table 2. Here, it can be seen that the math and reading scores for the average child in the sample at the fall of kindergarten were 23.155 and 29.372, respectively. Examining the coefficients on the two slope terms, it can be seen that the slopes of children's achievement trajectories were somewhat steeper between the fall and the spring of kindergarten, compared to the slopes of trajectories between the spring of kindergarten and fifth grade. The two slope terms in the unconditional growth models were correlated .80 for math and .66 for reading. Chi-square tests revealed that there was significant variability between individuals' achievement trajectories.

Given that children were heterogeneous in their academic trajectories, we next estimated conditional models of student achievement trajectories from the fall of kindergarten to the spring of fifth grade, with results for math displayed in Table 3 and findings for reading shown in Table 4. In Model 1 of Tables 3 and 4, an indicator of part-day kindergarten was used to predict the intercepts and slopes of children's achievement trajectories. Consistent with the descriptive analysis, there were no significant differences related to kindergarten program type at the fall of kindergarten, as indicated in Panel A. However, as shown in Panel B, the coefficients on the part-day indicator for the first slope in the piecewise growth models show that from fall to spring of the kindergarten year, the math and reading skills trajectories of children in part-day kindergarten were characterized by slightly slower rates of growth. More specifically, the math and reading skills of the average child in

Table 2

Kindergarten Program Type and Academic Trajectories From Fall of Kindergarten Through Spring of Fifth Grade

	Reading		Math	
	Coefficient	*SE*	Coefficient	*SE*
Intercept				
Intercept	29.372***	0.181	23.155***	0.170
Time 1[a]				
Intercept	2.803***	0.033	2.329***	0.042
Time 2[b]				
Intercept	1.672***	0.013	1.339***	0.014

[a] Rate of change during the kindergarten year.

[b] Rate of change from the spring of kindergarten until the spring of fifth grade.

$^{***}p < .001$.

part-day kindergarten grew by 2.189 and 2.632 per month, respectively, whereas the respective growth rates for full-day kindergarteners were 2.435 and 2.930. By the end of the kindergarten year, children in full-day kindergarten outscored part-day kindergarteners by one fifth of 1 *SD* in math and reading.

To consider whether the benefits of full-day kindergarten are sustained through the spring of fifth grade, we examined the coefficients on the second slope term in our piecewise growth models, which are found in Panel C of Model 1 of Tables 3 and 4. These suggest that the initial benefits of full-day kindergarten for the growth of achievement trajectories were not sustained beyond the kindergarten year. Indeed, the positive and significant coefficients on the part-day indicators for the second slope coefficients suggest that from the spring of kindergarten through fifth grade, the academic skills of children in part-day kindergarten grew at a slightly faster rate than did those of children in full-day kindergarten. Figure 1 illustrates the reading growth curve trajectories of children by kindergarten program type from the fall of kindergarten through the spring of third grade based on this first conditional model. The plot of children's math skills looks nearly identical. Here, it can be seen that although full-day kindergarten was associated with a small academic advantage across the kindergarten year, the achievement trajectories of children in part- and full-day kindergarten converge soon after children leave kindergarten. Indeed, Figure 1 shows that the advantage of full- versus part-day kindergarten has faded out by the spring of third grade, with the two groups diverging and the part-day kindergarteners pulling ahead. Though not shown here for parsimony, the divergence continues into the spring of fifth grade.

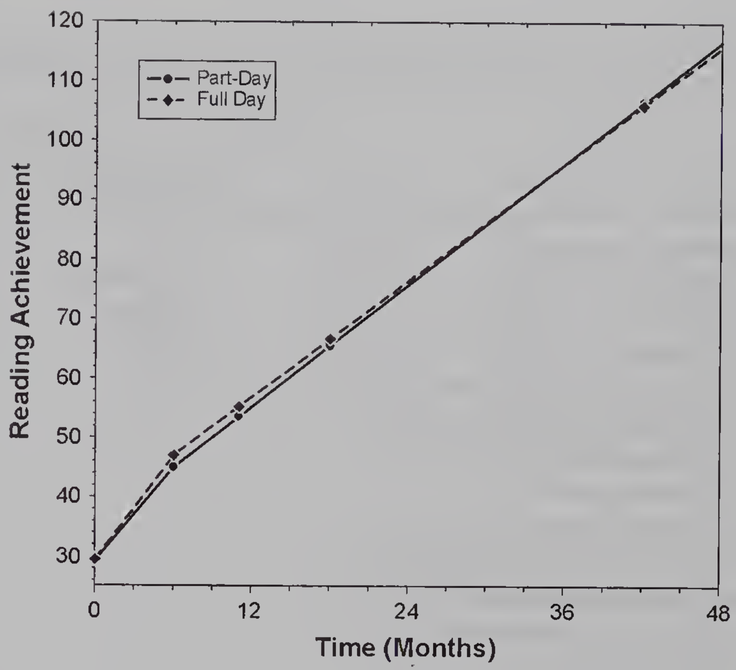


Figure 1. Kindergarten program type and reading achievement trajectories from fall of kindergarten to spring of third grade.

In Model 2 of Tables 3 and 4, measures of children's out-of-school contexts were introduced, including the quality of cognitive stimulation in the home environment, household poverty status, and the type and extent of nonparental care during kindergarten and the year before kindergarten entry. These factors were significantly related to children's achievement in the fall of kindergarten. More specifically, children who experienced high-quality home environments and nonparental care in the year before kindergarten (with the exception of Head Start enrollment), and those from nonpoor households tended to start school with higher levels of reading and math skills. The introduction of these variables as predictors of the intercept revealed a significant difference in achievement between children in part- and full-day kindergarten that favored full-day kindergarten students, though this association was only marginally significant for math achievement.

The inclusion of this important set of variables as predictors of the kindergarten slope term resulted in a 26% and 19% reduction in the association between kindergarten program type and children's math and reading skills trajectories, respectively, during the kindergarten year, highlighting the importance of these contexts for children's development. Together, cognitive stimulation in the home environment and the poverty status of the child's household were responsible for a modest decrease in the magnitude of the part-day coefficient. More specifically, the math and reading skills of children from impoverished households grew at slower rates than did those of children in households above the federal poverty

Table 3

Kindergarten Program Type and Math Trajectories From Fall of Kindergarten to Spring of Fifth Grade

	Model 1		Model 2		Model 3	
	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*
Panel A. Intercept						
Intercept	22.987***	0.240	23.241***	0.174	23.553***	0.147
Part day	0.358	0.319	−0.416†	0.246	−0.706**	0.203
Poverty			−3.604***	0.203	−0.885***	0.195
Home learning environment			3.997***	0.245	2.165***	0.198
Hours per week in child care[a]			−0.012*	0.005	0.000	0.005
Head Start[a]			−1.196**	0.371	−0.421	0.345
Center care[a]			3.365***	0.295	1.676***	0.266
Relative home-based care[a]			0.965*	0.380	0.361	0.355
Nonrelative home-based care[a]			2.857***	0.428	1.106**	0.406
Multiple types of care[a]			1.934***	0.498	0.782†	0.459
Hours per week in child care[b]			−0.015†	0.008	−0.001	0.008
Center care[b]			0.392	0.278	−0.075	0.247
Relative home-based care[b]			−0.980*	0.290	−0.330	0.253
Nonrelative home-based care[b]			0.489	0.360	0.144	0.318
Multiple types of care[b]			−0.064	0.559	−0.054	0.483
Age of entry					0.564***	0.022
Changed teacher during kindergarten					−0.811*	0.339
Male					0.054	0.142
Race						
African American					−1.870***	0.228
Hispanic					−1.800***	0.264
Non-Hispanic Asian					1.773**	0.506
Native American					−2.801***	0.453
Multiracial					1.006*	0.434
Low birth weight					1.796***	0.239
Maternal age at first birth (years)					0.142***	0.019
Highest parental education						
High school degree					0.917**	0.269
Vocational/technical program					2.061***	0.281
Bachelor's degree					3.944***	0.333
Advanced degree					5.871***	0.398
Marital status						
Divorced/separated					−0.642**	0.209
Widowed					−1.047	0.732
Never married					−0.536*	0.233
Adoptive/no biological dad					−1.501**	0.420
Non-English home language					−0.752*	0.323
Number of children in household					−0.259***	0.067
Midwest					0.184	0.285
West					0.895**	0.306
South					0.049	0.295
Private					1.434***	0.296
Central city					−0.261	0.210
Rural					−1.222***	0.230
Panel B. Time 1[c]						
Intercept	2.435***	0.048	2.461	0.045	2.483***	0.042
Part day	−0.246***	0.060	−0.183***	0.048	−0.153**	0.047
Poverty			−0.072†	0.037	−0.077*	0.039
Home learning environment			0.112**	0.038	0.106**	0.032

Table 3
Continued

	Model 1		Model 2		Model 3	
	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*
Hours per week in child care[b]			−0.001	0.001	−0.000	0.001
Center care[b]			−0.028	0.044	−0.015	0.041
Relative home-based care[b]			−0.006	0.040	−0.035	0.039
Nonrelative home-based care[b]			0.057	0.053	0.062	0.050
Multiple types of care[b]			0.071	0.091	0.017	0.090
Age of entry					0.027***	0.003
Changed teacher during kindergarten					−0.101	0.120
Male					0.072**	0.025
Race						
African American					−0.447***	0.046
Hispanic					−0.132**	0.048
Non-Hispanic Asian					−0.112	0.073
Native American					−0.087	0.079
Multiracial					−0.191*	0.079
Low birth weight					−0.120*	0.047
Maternal age at first birth (years)					0.002	0.003
Highest parental education						
High school degree					0.075	0.048
Vocational/technical program					0.201***	0.052
Bachelor's degree					0.280***	0.062
Advanced degree					0.414***	0.071
Marital status						
Divorced/separated					−0.013	0.038
Widowed					0.026	0.132
Never married					−0.074	0.047
Adoptive/no biological dad					−0.097	0.083
Non-English home language					−0.068	0.055
Number of children in household					−0.002	0.011
Midwest					0.252***	0.055
West					0.148*	0.060
South					0.324***	0.061
Private					−0.015	0.054
Central city					−0.048	0.041
Rural					−0.111*	0.051
Panel C. Time 2[d]						
Intercept	1.310***	0.015	1.316***	0.014	1.328***	0.014
Part day	0.065***	0.009	0.059***	0.008	0.038***	0.008
Poverty			−0.126***	0.009	−0.067***	0.009
Home learning environment			0.037*	0.012	0.007	0.011
Age of entry					−0.004***	0.001
Changed teacher during kindergarten					−0.031	0.020
Male					0.061***	0.007
Race						
African American					−0.088***	0.012
Hispanic					−0.006	0.011
Non-Hispanic Asian					0.040**	0.015
Native American					−0.052*	0.021
Multiracial					−0.018	0.021
Low birth weight					−0.028*	0.012
Maternal age at first birth (years)					0.003**	0.001

(*Continued*)

Table 3
Continued

	Model 1		Model 2		Model 3	
	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*
Highest parental education						
High school degree					0.039**	0.014
Vocational/technical program					0.099***	0.012
Bachelor's degree					0.138***	0.015
Advanced degree					0.158***	0.016
Marital status						
Divorced/separated					0.003	0.011
Widowed					−0.016	0.037
Never married					−0.012	0.013
Adoptive/no biological dad					−0.006	0.023
Non-English home language					0.003	0.013
Number of children in household					−0.007*	0.003
Midwest					−0.024*	0.011
West					−0.017	0.013
South					−0.010	0.013
Private					−0.041***	0.010
Central city					0.005	0.009
Rural					−0.027*	0.012

[a] Child-care characteristics at prekindergarten.
[b] Child-care characteristics at kindergarten.
[c] Rate of change during the kindergarten year.
[d] Rate of change from the spring of kindergarten until the spring of fifth grade.
[†] $p < .10$. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

threshold, whereas the achievement trajectories of children from more stimulating home environments increased at steeper rates during the kindergarten year. After taking these differences into account, the disadvantage of attending part-day kindergarten appeared modestly smaller. Contrary to our hypothesis, kindergarten child-care characteristics did not emerge as significant predictors of achievement trajectories during the kindergarten year.

The introduction of out-of-school contexts also slightly decreased the part-day kindergarten advantage for children's achievement trajectories from the spring of kindergarten to the spring of fifth grade. Indeed, the coefficient on the part-day kindergarten indicator for the second slope of children's math and reading trajectories fell by about 9% with the introduction of the additional measures into the analysis. Again, cognitive stimulation in the home environment and family poverty status were responsible for this reduction. This suggests that the convergence of the achievement trajectories of children in part- and full-day kindergarten may in part be explained by the lower levels of poverty and the more stimulating home environments of children enrolled in part-day, as opposed to full-day kindergarten programs. Based on this second set of conditional models, it can be anticipated that by the end of the kindergarten year, children in full-day kindergarten outscored part-day kindergarteners by 0.11 of 1 *SD* for math and 0.15 of 1 *SD* for reading, which is one half to one quarter of a reduction in the expected gain based on the first set of conditional models.

In Model 3 of Tables 3 and 4, an extensive set of potentially confounding factors were added as predictors of children's achievement trajectories, with child and family characteristics added to the Level 2 equations and school characteristics added at Level 3. The introduction of this comprehensive set of control variables resulted in slight increases in the association between kindergarten program type and academic achievement at the fall of kindergarten. More importantly, however, it led to further reduction of the benefit of full-day kindergarten for the growth of academic skills during kindergarten. Indeed, the part-day kindergarten coefficients on the kindergarten slope terms in the prior conditional models decreased 16% and 8% for math and reading, respectively. The quality of children's home environment and their household poverty status continued to be significantly associated with achievement trajectories during kindergarten, though the sizes of these coefficients dropped with the inclusion of the more comprehensive set of

Table 4

Kindergarten Program Type and Reading Trajectories From Fall of Kindergarten to Spring of Fifth Grade

	Model 1		Model 2		Model 3	
	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*
Panel A. Intercept						
Intercept	29.455***	0.259	29.714***	0.200	29.962***	0.183
Part day	−0.216	0.343	−1.008*	0.285	−1.081***	0.259
Poverty			−3.449***	0.222	−0.829**	0.228
Home learning environment			4.576***	0.282	2.697***	0.234
Hours per week in child care[a]			−0.009	0.007	−0.002	0.007
Head Start[a]			−1.568**	0.433	−0.992*	0.417
Center care[a]			3.758***	0.352	1.812***	0.337
Relative home-based care[a]			1.047*	0.479	0.419	0.470
Nonrelative home-based care[a]			2.607***	0.542	0.797	0.513
Multiple types of care[a]			2.155**	0.631	0.802	0.602
Hours per week in child care[b]			−0.003	0.009	0.005	0.010
Center care[b]			−0.061	0.313	−0.631*	0.288
Relative home-based care[b]			−1.304***	0.330	−0.863**	0.311
Nonrelative home-based care[b]			0.006	0.459	−0.237	0.445
Multiple types of care[b]			−0.720	0.614	−0.626	0.569
Age of entry					0.444***	0.025
Changed teacher during kindergarten					−0.527	0.401
Male					−1.449***	0.162
Race						
African American					0.069	0.333
Hispanic					−0.945**	0.322
Non-Hispanic Asian					3.409***	0.704
Native American					−2.029**	0.587
Multiracial					0.287	0.667
Low birth weight					−1.427***	0.283
Maternal age at first birth (years)					0.153***	0.021
Highest parental education						
High school degree					1.173***	0.278
Vocational/technical program					2.235***	0.285
Bachelor's degree					4.000***	0.355
Advanced degree					6.453***	0.466
Marital status						
Divorced/separated					−0.870***	0.227
Widowed					−0.062	1.046
Never married					−0.675*	0.283
Adoptive/no biological dad					−1.116*	0.458
Non-English home language					−1.123*	0.488
Number of children in household					−0.729***	0.077
Midwest					−0.003	0.341
West					1.255**	0.383
South					0.346	0.316
Private					1.661***	0.362
Central city					−0.576*	0.247
Rural					−1.339***	0.295
Panel B. Time 1[c]						
Intercept	2.930***	0.047	2.970***	0.043	3.030***	0.042
Part day	−0.298***	0.073	−0.243***	0.062	−0.243***	0.063
Poverty			−0.290***	0.046	−0.178***	0.049
Home learning environment			0.328***	0.050	0.253***	0.045

(*Continued*)

Table 4
Continued

	Model 1		Model 2		Model 3	
	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*
Hours per week in child care[b]			−0.003	0.002	−0.001	0.002
Center care[b]			−0.032	0.059	−0.093	0.057
Relative home-based care[b]			−0.020	0.058	−0.059	0.055
Nonrelative home-based care[b]			0.089	0.064	0.017	0.061
Multiple types of care[b]			0.052	0.115	−0.024	0.112
Age of entry					0.031***	0.004
Changed teacher during kindergarten					−0.272*	0.105
Male					−0.257***	0.033
Race						
African American					−0.313***	0.069
Hispanic					−0.114†	0.066
Non-Hispanic Asian					0.368***	0.099
Native American					−0.216†	0.121
Multiracial					−0.011	0.100
Low birth weight					−0.171*	0.067
Maternal age at first birth (years)					0.004	0.004
Highest parental education						
High school degree					0.154**	0.057
Vocational/technical program					0.331***	0.064
Bachelor's degree					0.445***	0.070
Advanced degree					0.624***	0.088
Marital status						
Divorced/separated					−0.037	0.050
Widowed					0.070	0.178
Never married					−0.231***	0.055
Adoptive/no biological dad					−0.219*	0.105
Non-English home language					−0.120†	0.070
Number of children in household					−0.062***	0.015
Midwest					0.095	0.079
West					0.184*	0.090
South					0.234**	0.081
Private					0.136	0.088
Central city					−0.029	0.059
Rural					−0.124	0.076
Panel C. Time 2[d]						
Intercept	1.639***	0.014	1.646***	0.014	1.655***	0.014
Part day	0.073***	0.011	0.066***	0.010	0.055***	0.010
Poverty			−0.149***	−0.012	0.078***	0.012
Home learning environment			0.042***	0.009	−0.005	0.010
Age of entry					−0.002*	0.001
Changed teacher during kindergarten					−0.009	0.023
Male					−0.015	0.010
Race						
African American					−0.093***	0.014
Hispanic					−0.026†	0.015
Non-Hispanic Asian					−0.080***	0.020
Native American					−0.097***	0.025
Multiracial					−0.043†	0.023
Low birth weight					0.005	0.016
Maternal age at first birth (years)					0.002**	0.001

Table 4
Continued

	Model 1		Model 2		Model 3	
	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*
Highest parental education						
High school degree					0.087***	0.017
Vocational/technical program					0.139***	0.016
Bachelor's degree					0.191***	0.018
Advanced degree					0.207***	0.020
Marital status						
Divorced/separated					0.009	0.012
Widowed					−0.034	0.041
Never married					0.002	0.013
Adoptive/no biological dad					0.034	0.023
Non-English home language					−0.025†	0.014
Number of children in household					−0.013***	0.003
Midwest					0.000	0.014
West					−0.038*	0.015
South					−0.016	0.014
Private					0.004	0.013
Central city					−0.006	0.010
Rural					−0.028†	0.014

[a] Child-care characteristics at prekindergarten.
[b] Child-care characteristics at kindergarten.
[c] Rate of change during the kindergarten year.
[d] Rate of change from the spring of kindergarten until the spring of fifth grade.
† $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

covariates. The introduction of this extensive set of covariates in Model 3 of Tables 3 and 4 further attenuated the association between kindergarten program type and achievement trajectories from the spring of kindergarten to fifth grade as well. Indeed, compared to Model 2, 36% of the part-day growth advantage in math and 17% of the part-day growth advantage for reading were explained by this set of covariates. Although child poverty continued to be significantly related to the second slope term, cognitive stimulation in the home environment was no longer a significant predictor. Thus, it appears that differences in poverty status, not cognitive stimulation in the home environment, may be responsible for explaining the part-day kindergarten advantage in academic growth from the spring of kindergarten through fifth grade.

Discussion

Full-day kindergarten experiences are becoming increasingly common for children across the United States. Yet, there are notable gaps in the literature regarding whether full-day kindergarten attendance is advantageous for children's development. This is the first study to take a developmental approach to examining the implications of full- versus part-day kindergarten for children's learning trajectories. Specifically, this investigation recognizes both the heterogeneity of children's achievement trajectories and the complex configuration of child, family, and nonparental care factors that affect their developmental trajectories. In doing so, this study provides important insights about the size and persistence of the full-day kindergarten advantage and strengthens our understanding of when and why the full-day kindergarten advantage seems to fade out soon after the end of the kindergarten year.

Full- Versus Part-Day Kindergarteners: A Descriptive Portrait

Our descriptive portrayal of a nationally representative sample of kindergarteners revealed several significant, but generally modest, differences in out-of-school dimensions and individual characteristics of children in part- versus full-day kindergarten programs. Among the most notable difference is the racial composition of the two groups of children. More specifically, a greater proportion of children in part-day kindergarten were of Hispanic, non-Hispanic White, Asian, and multiracial backgrounds, and a smaller proportion were of non-Hispanic Black and Native American race/ethnicity, in comparison to full-day kindergarteners. The largest difference when it came to racial composition was in the proportion of children of non-Hispanic Black ethnicity in each group. This proportion was three times higher among full-day kindergarteners (22%) when compared to part-day kindergarten students (7%). On the whole, the bivariate analysis suggests that part-day kindergarteners were more advantaged socioeconomically; however, the multivariate analyses that controlled for these differences surprisingly revealed higher levels of academic achievement at the fall of kindergarten among children enrolled in *full*-day kindergarten. Future research that more carefully models families' selection of full- versus part-day programs for their children is needed to untangle the more complex story detected here.

Benefits of Full-Day Kindergarten Attendance During the Transition to School

The results of this study concur with the existing body of research that has largely shown significant academic benefits of full- versus part-day kindergarten programs (e.g., Cannon, Jacknowitz, & Painter, 2006; Clark & Kirk, 2000; Gullo, 2000; Kaplan, 2002; Lee et al., 2006; Walston & West, 2004; Weiss & Offenberg, 2002). Unlike prior studies, however, the current investigation linked kindergarten program type not only to levels of academic achievement but also to individual growth trajectories of students' math and reading skills during the kindergarten year using data from a nationally representative study. Without controls for important child, family, and nonparental care characteristics, full-day kindergarten students' achievement grew at a rate that was 0.246 points per month faster in math and 0.298 points per month faster in reading. This amounted to a modest advantage in math and reading achievement, which was roughly one fifth of 1 *SD*. After an extensive set of child, family, school, and nonparental care characteristics were introduced to the growth models, the magnitude of the full-day kindergarten advantage dropped in half for math and by one quarter for reading, resulting in estimates of the full-day kindergarten advantage that are quite small and are similar to those found by Cannon et al. (2006) with these same data.

The modest nature of the full-day advantage may be attributed to slight differences in the amount and type of instruction taking place in kindergarten classrooms. Although teachers participating in the ECLS-K report that full-day kindergarteners spent roughly two times as much time in school as half-day kindergarteners, they did not receive two times as much instruction (Xue, Burkam, & Lee, 2002). For example, full-day kindergarten teachers spend 5.5 hr per week in reading and language arts and 3.7 hr in mathematics, whereas part-day kindergarten teachers spend 4.3 hr in reading and language arts and 2.5 hr in mathematics (Xue et al., 2002). Furthermore, full- and part-day classes are organized in similar ways, in terms of the proportion of time spent in whole class, small group, individual, and child-selected activities (Walston & West, 2002). Additional research on the heterogeneity of instructional practices and of teacher – child interactions that occur within part- and full-day kindergarten programs may help identify whether characteristics of kindergarten beyond program type are more successful in promoting student achievement.

Fading Benefits of Full-Day Kindergarten Attendance During Early Elementary School

Perhaps the most troubling finding of the present study is that the academic benefits of full-day kindergarten subside soon after children leave kindergarten. This is congruent with past research that has found that the academic benefits of full-day kindergarten are relatively short lived (Karweit, 1987; Ohio State Legislative Office of Education, 1997; Weiss & Offenberg, 2002). Unlike prior research, this developmental approach paints a more nuanced understanding of the fade-out of the full-day kindergarten advantage, both by providing a more precise estimate of when the fade-out occurs and by shedding light on factors that are partially responsible for explaining why the trajectories of children in part- and full-day kindergarten converge. The fade-out occurs as the trajectories of children in part-day kindergarten grow at a steeper rate from the spring of kindergarten to the spring of fifth grade. Our models suggest that the advantage of full- versus part-day kindergarten fades out approximately 36 months after the spring of kindergarten assessment, or in the spring of third grade. Furthermore, the fade-out appears to be attributed to differences in the children and families that attend part- and full-day kindergarten, as well as school characteristics associated with kindergarten program type. These differences explained 42% of the part-day kindergarten growth advantage in math from the spring of kindergarten through the spring of fifth grade and 25% of the part-day growth advantage for reading. Thus, it seems that the greater selection of children from, for example, economically disadvantaged households and a non-Hispanic Black ethnic background contributes to the fade-out of the full-day kindergarten advantage after the end of the kindergarten year. It is important to recognize, however, that this study is not able to rule out the threat of omitted variable bias when examining the fade-out of the full-day kindergarten advantage during the early elementary school years. Indeed, we are not able to rule out the possibility that it may be an unmeasured characteristic of parents or children that jointly explain families' home environment quality and economic status, which seem to be important in explaining the fade-out of the full-day kindergarten advantage.

Furthermore, this study fell short of entirely explaining why the growth rates of children attending part-day kindergarten programs tend to catch up to their full-day kindergarten counterparts so quickly. Other factors, such as subsequent schooling experiences, may be at play. Research on Head Start programs, the federally funded early childhood education intervention for economically disadvantaged children, lends support for this hypothesis. Lee and Loeb (1995) as well as Currie and Thomas (2000) found that differences in the schooling experiences subsequent to Head Start participation are important for understanding why Head Start effects seem to fade out over time. For example, using data from the National Education Longitudinal Study of 1988, Lee and Loeb found that children who attended Head Start were educated in schools of significantly lower quality, defined by social composition, academic rigor, safety, and social relations, when compared to counterparts who did not attend preschool and especially when compared to peers who attended other preschools. Similarly, more socioeconomically disadvantaged full-day kindergarten students may be attending schools of lower quality compared to the quality of schools attended by part-day kindergarteners. In other words, differences in children's later schooling experiences may erode the benefits associated with both Head Start and full-day kindergarten participation.

Policy Implications

Though it is impossible to reduce the threats of selection or omitted variable bias entirely with nonexperimental data, the results of this study suggest that the shift from part- to full-day kindergarten programs occurring across the United States may have positive implications for students' learning trajectories, at least in the short run. Here, we find modest but meaningful differences in the achievement trajectories of full- and part-day kindergarten students that favor full-day kindergarten programs. What is perhaps most concerning is the immediate nature of these benefits, which raises questions about whether the cost of full-day as opposed to part-day kindergarten programs is associated with sufficient benefits for children and society.

It is difficult based on the existing literature to determine the cost of full-day kindergarten programs. Cost estimates range substantially, even within the same state. According to Weiss and Offenberg (2002), who have conducted a cost–benefit analysis of full-day kindergarten in the School District of Philadelphia, the cost of full-day kindergarten is approximately $2,300 per year per child. Yet, the Pennsylvania Partnerships for Children (2000) estimated the median instructional expense of full-day kindergarten per pupil to be $5,216. A comprehensive cost–benefit analysis would be required to evaluate the relative effectiveness of public investments in full-day kindergarten as opposed to investment in other programs for children. To carefully conduct such an analysis, it would be important to look at other domains of student functioning in addition to academic achievement, such as measures of socioemotional functioning and physical health and well-being. Of course, the other rather obvious benefit of full-day kindergarten that must be taken into account in a systematic analysis is its linkage with parental employment. Cannon et al. (2006) found that mothers are more likely to work full time in the kindergarten year if their children attend full-day kindergarten. Increases in parental employment benefit society with higher tax revenue and can boost families' household income. It may be that these benefits end up being more important in making the case for full-day kindergarten in the policy arena than are achievement gains made by full-day kindergarteners. Although the present study provides useful information about associations between kindergarten program type and two domains of achievement, a more systematic accounting of the costs and benefits of the program, similar to what has been conducted for the Perry Preschool Project (Barnett, 1996; Belfield, Nores, Barnett, & Schweinhart, 2006), is necessary to evaluate public investments in full-day kindergarten.

Conclusions

In sum, this study advances the literature on full- versus part-day kindergarten attendance by situating such programs in a developmental context. Unlike prior studies in this research area, we recognized individual differences in children's achievement trajectories as well as the constellation of child characteristics, family factors, and nonparental care experiences that add nuance and texture to children's lives. In doing so, this study moved beyond static views of children's achievement and revealed that attending full-day kindergarten yielded a modest benefit to children's academic trajectories over time. Furthermore, our developmental approach highlighted the importance of considering not only whether children attending full-day programs were able to sustain these benefits but also when children enrolled in full-day programs no longer maintained this advantage. Here, we pinpointed that academic benefits associated with full-day kindergarten program attendance faded out by the end of the primary grades. Last, our developmental view brought children's background characteristics and experiences into the "foreground" of understanding differences between children attending full- versus part-day kindergarten programs. In focusing on the multifaceted nature of children's lives, we found that child and family characteristics played noteworthy roles in why full-day benefits exist and in why these advantages fade relatively quickly. Incorporating a developmental perspective into future studies of policies relevant to young children's educational experiences may similarly prove fruitful.

References

Acock, A. C. (2005). Working with missing values. *Journal of Marriage and Family, 67*, 1012 – 1028.

Alexander, K. L., Entwisle, D. R., & Dauber, S. L. (1993). First-grade classroom behavior: Its short- and long-term consequences for school performance. *Child Development, 64*, 801 – 814.

Allison, P. D. (1990). Change scores as dependent variables in regression analysis. In C. C. Clogg (Ed.), *Sociological methodology* (pp. 93 – 114). Oxford, England: Basil Blackwell.

Barnett, W. S. (1996). *Lives in balance: Age 27 benefit-cost analysis of the High/Scope Perry Preschool Program* (Monographs of the High/Scope Educational Research Foundation, 11). Ypsilanti, MI: High/Scope Press.

Belfield, C. R., Nores, M., Barnett, S., & Schweinhart, L. (2006). The High/Scope Perry Preschool Program: Cost-benefit analysis using data from the age-40 follow-up. *Journal of Human Resources, 41*, 162 – 190.

Bogenschneider, K., & Steinberg, L. (1994). Maternal employment and adolescents' academic achievement: A developmental analysis. *Sociology of Education, 67*, 60 – 77.

Bradley, R. H., Corwyn, R. F., Burchinal, M., McAdoo, H. P., & Garcia Coll, C. (2001). The home environment of children in the United States Part II: Relations with behavioral development through age thirteen. *Child Development, 72*, 1868 – 1886.

Broberg, A. G., Wessels, H., Lamb, M. E., & Hwang, C. P. (1997). Effects of day care on the development of cognitive abilities in 8-year-olds: A longitudinal study. *Developmental Psychology, 33*, 62 – 69.

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Cambridge, MA: Harvard University Press.

Bronfenbrenner, U., & Morris, P. A. (2006). The bioecological model of human development. In R. Lerner (Vol. Ed.), *Handbook of child psychology: Theoretical models of human development* (Vol. 1, 6th ed., pp. 793 – 893). New York: John Wiley.

Brooks-Gunn, J., Han W., & Waldfogel, J. (2002). Maternal employment and child cognitive outcomes in the first three years of life: The NICHD study of early child care. *Child Development, 73*, 1052 – 1072.

Brooks-Gunn, J., Klebanov, P. K., & Liaw, F. R. (1995). The learning, physical, and emotional environment of the home in the context of poverty: The Infant Health and Development Program. *Children and Youth Services Review, 17*, 251 – 276.

Burchinal, M., & Appelbaum, M. I. (1991). Estimating individual developmental functions: Methods and their assumptions. *Child Development, 62*, 23 – 43.

Burchinal, M., Nelson, L., & Poe, M. (2006). Growth curve analysis: An introduction to various methods for analyzing longitudinal data. *Monographs of the Society for Research in Child Development, 71*(3).

Burriss, K. G. (2000). All-day kindergarten. *Childhood Education, 76*, 228 – 231.

Caldwell, B. M., & Bradley, R. H. (1979). *Home observation for measurement of the environment*. Little Rock: Center for Child Development and Education, University of Arkansas.

Cannon, J. S., Jacknowitz, A., & Painter, G. (2006). Is full better than half? Examining the longitudinal effects of full-day kindergarten attendance. *Journal of Policy Analysis and Management, 25*, 299 – 321.

Caughy, M., DiPietro, J., & Strobino, D. (1994). Day-care participation as a protective factor in the cognitive development of low-income children. *Child development, 65*, 457 – 471.

Clark, P., & Kirk, E. (2000). All-day kindergarten. *Childhood Education, 76*, 228 – 231.

Coleman, J. S. (1966). *Equality of educational opportunity.* Washington, DC: U.S. Department of Health, Education, and Welfare, Office of Education.

Cryan, J., Sheehan, R., Wiechel, J., & Bandy-Hedden, I. (1992). Success outcomes of full-day kindergarten: More positive behavior and increased achievement in the years after. *Early Childhood Research Quarterly, 7*, 187 – 203.

Currie, J., & Thomas, D. (2000). School quality and the longer-term effects of Head Start. *Journal of Human Resources, 35*, 755 – 774.

Dearing, E., McCartney, K., & Taylor, B. A. (2001). Change in family income-to-needs matters more for children with less. *Child Development, 72*, 1779 – 1793.

Duncan, G. J., & Brooks-Gunn, J. (1997). *Consequences of growing up poor.* New York: Russell Sage Foundation.

Duncan, G. J., Brooks-Gunn, J., Yeung, W. J., & Smith, J. R. (1998). How much does childhood poverty affect the life chances of children? *American Sociological Review, 63,* 406–423.

Dunn, L. M., & Dunn, L. M. (1981). *PPVT–Revised manual.* Circle Pines, MN: American Guidance Service.

Dunn, L. M., & Markwardt, F. C. (1970). *Peabody individual achievement test manual.* Circle Pines, MN: American Guidance Service.

Elicker, J., & Mathur, S. (1997). What do they do all day? Comprehensive evaluation of a full-day kindergarten. *Early Childhood Research Quarterly, 7,* 187–204.

Entwisle, D. R., Alexander, K. L., Cadigan, D., & Pallas, A. M. (1987). Kindergarten experience: Cognitive effects or Socialization? *American Educational Research Journal, 24,* 337–364.

Entwisle, D. R., Alexander, K. L., & Olson, L. S. (1997). *Children, schools, and inequality.* Boulder, CO: Westview Press.

Evans, G. W., & English, K. (2002). The environment of poverty: Multiple stressor exposure, psychophysiological stress, and socioemotional adjustment. *Child Development, 73,* 1238–1248.

Field, T. (1991). Quality infant day-care and grade school behavior and performance. *Child Development, 62,* 863–870.

Foster, E. M., & Kalil, A. (2005). Developmental psychology and public policy: Progress and prospects. *Developmental Psychology, 41,* 827–832.

Gullo, D. F. (2000). The long-term educational effects of half-day vs. full-day kindergarten. *Early Childhood Development and Care, 160,* 17–24.

Gullo, D. F., Bersani, C. U., Clements, D. H., & Bayless, K. M. (1986). A comparative study of "all-day," "alternate-day," and "half-day" kindergarten schedules: Effects on achievement and classroom social behaviors. *Journal of Research in Childhood Education, 1,* 87–94.

Harvey, E. (1999). Short-term and long-term effects of early parental employment on children of the National Longitudinal Survey of Youth. *Child Development, 35,* 445–459.

Huber, P. J. (1967). The behavior or maximum likelihood estimates under non-standard conditions. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 4,* 221–233.

Kaplan, D. (2002). Methodological advances in the analysis of individual growth with relevance to education policy. *Peabody Journal of Education, 77,* 189–215.

Karweit, N. (1987). *Full or half day kindergarten—Does it matter?* (Report No. 11). Baltimore: Center for Research on Elementary and Middle Schools, The Johns Hopkins University.

Karweit, N. (1992). The kindergarten experience. *Educational Leadership, 49,* 82–86.

Lee, V. E., Burkam, D. T., Ready, D. D., Honigman, J., & Meisels, S. J. (2006). Full-day vs. half-day kindergarten: In which program do children learn more? *American Journal of Education, 112,* 163–208.

Lee, V. E., & Loeb, S. (1995). Where do Head Start attendees end up? One reason why preschool effects fade out. *Educational Evaluation and Policy Analysis, 17,* 62–82.

Markwardt, F. C. (1989). *Peabody Individual Achievement test–Revised.* Circle Pines, MN: American Guidance Service.

McCoach, D. B., O'Connell, A. A., Reis, S. M., & Levitt (2006). Growing readers: A hierarchical linear model of children's reading growth during the first 2 years of school. *Journal of Educational Psychology, 98,* 14–28.

McLoyd, V. C. (1998). Socioeconomic disadvantage and child development. *American Psychologist, 53,* 185–204.

Morrison, F. J., Bachman, H. J., & Connor, C. M. (2005). *Improving literacy in America: Guidelines from research.* New Haven, CT: Yale University Press.

National Center for Children in Poverty. (2006). *Basic facts about low-income children: Birth to age 6.* New York: National Center for Children in Poverty, Columbia University, Mailman School of Public Health.

National Institute of Child Health and Human Development Early Child Care Research Network. (2000). The relation of child care to cognitive and language development. *Child Development, 71,* 960–980.

National Institute of Child Health and Human Development Early Child Care Research Network. (2002). Early child care and children's development prior to school entry. *American Educational Research Journal, 39,* 133–164.

National Research Council and Institute of Medicine. (2000). From neurons to neighborhoods: The science of early childhood development. Committee on Integrating the Science of Early Childhood Development. In J. P. Shonkoff & D. A. Phillips (Eds.), *Board on Children, Youth, and Families, Commission on Behavioral and Social Sciences and Education.* Washington, DC: National Academy Press.

Ohio State Legislative Office of Education. (1997). An overview of full-day kindergarten. Columbus: Ohio State Legislative Office of Education.

Patterson, C. J., Kupersmidt, J. B., & Vaden, N. A. (1990). Income level, gender, ethnicity, and household composition as predictors of children's school-based competence. *Child Development, 61,* 485–494.

The Pennsylvania Partnerships for Children (2000). *Learning to learn: Full-day kindergarten for at-risk kids.* Harrisburg: Pennsylvania Partnerships for Children.

Pianta, R., Rimm-Kaufman, S. E., & Cox, M. J. (1999). Introduction: An ecological approach to kindergarten transition. In R. C. Pianta & M. J. Cox (Eds.), *The transition to kindergarten* (pp. 3–12). Baltimore: Brookes Publishing.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.) Thousand Oaks, CA: Sage.

Raudenbush, S. W., Bryk, A., Cheong, Y. F., Congdon, R., & du Toit, M. (2004). *HLM 6: Hierarchical linear and nonlinear modeling.* Lincolnwood, IL: Scientific Software International.

Ready, D. D., LoGerfo, L. F., Burkam, D. T., & Lee, V. E. (2005). Explaining girls' advantage in kindergarten literacy learning: Do classroom behaviors make a difference? *Elementary School Journal, 106*, 21–38.

Rogosa, D., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin, 92*, 726–748.

Royston, P. (2004). Multiple imputation of missing values. *Stata Journal, 4*, 227–241.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: John Wiley & Sons.

Rutter, M. (1981). Social-emotional consequences of day care for preschool children. *American Journal of Orthopsychiatry, 51*, 4–28.

Shonkoff, J. & Phillips, D. A. (Eds.). (2000). *From neurons to neighborhoods: The science of early child development*. Washington, DC: National Academy of Sciences.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford University Press.

Smith, J. R., Brooks-Gunn, J., & Klebanov, P. K. (1997). Consequences of living in poverty for young children's cognitive and verbal ability and early school achievement. In G. J. Duncan & J. Brooks-Gunn (Eds.), *Consequences of growing up poor* (pp. 132–189). New York: Russell Sage Foundation.

U.S. Department of Education, National Center for Education Statistics. (2001). *Early Childhood Longitudinal Study, Kindergarten Class of 1998–99: Users manual for the ECLS-K base year public-use data files and electronic codebook* (NCES 2001-029). Washington, DC: Author.

U.S. Department of Education, National Center for Education Statistics. (2005). *Early childhood longitudinal study, kindergarten class of 1998–99: Psychometric report for the fifth grade* (NCES 2006-036). Washington, DC: Author.

Votruba-Drzal, E. (2003). Income changes and cognitive stimulation in young children's home environments. *Journal of Marriage and Family, 65*, 341–355.

Votruba-Drzal, E. (2006). Economic disparities in middle childhood development: Does income matter? *Developmental Psychology, 42*, 1154–1167.

Votruba-Drzal, E., Coley, R. L., & Chase-Lansdale, P. L. (2004). Child care quality and low-income children's development: For who is child care most important? *Child Development, 75*, 296–312.

Walston, J., & West, J. (2002, June). *Classroom organization and curriculum differences between full-day and part-day kindergarten programs in the nation's public schools*. Paper presented at the American Educational Research Association's Annual Meeting, New Orleans, LA.

Walston, J., & West, J. (2004). *Full-day and half-day kindergarten in the United States: Findings from the Early Childhood Longitudinal Study, kindergarten class of 1998–1999* (NCES-2004-078). Washington, DC: U.S. Government Printing Office; U.S. Department of Education. National Center for Education Statistics.

Weiss, A., & Offenberg, R. M. (2002, April). *Enhancing urban children's early success in school: The power of full-day kindergarten*. Paper presented at the Annual Conference of the American Educational Research Association Meeting, New Orleans, LA.

Widaman, K. F. (2006). Missing data: What to do with or without them. *Monographs of the Society for Research in Child Development, 71*, 42–64.

Willett, J. B. (1989). Questions and answers to the measurement of change. *Review of Research in Education, 15*, 345–422.

Woodcock, R. W., & Mather, N. (1989, 1990). WJ-R tests of achievement: examiner's manual. In R. W. Woodcock & M. B. Johnson, *Woodcock-Johnson Psycho-Educational Battery–Revised*. Itasca, IL: Riverside Publishing.

Xue, Y., Burkam, D. T., & Lee, V. E. (2002, April). *Unpacking full-day/half-day kindergarten: Is full-day kindergarten double-dose of half-day kindergarten?* Paper presented at the American Educational Research Association's Annual Meeting, New Orleans, LA.

Child Development, July/August 2008, Volume 79, Number 4, Pages 979 – 1000

Novel Noun and Verb Learning in Chinese-, English-, and Japanese-Speaking Children

Mutsumi Imai
Keio University

Lianjing Li
Peking University

Etsuko Haryu
University of Tokyo

Hiroyuki Okada
Tamagawa University

Kathy Hirsh-Pasek
Temple University

Roberta Michnick Golinkoff
University of Delaware

Jun Shigematsu
Keio University

When can children speaking Japanese, English, or Chinese map and extend novel nouns and verbs? Across 6 studies, 3- and 5-year-old children in all 3 languages map and extend novel nouns more readily than novel verbs. This finding prevails even in languages like Chinese and Japanese that are assumed to be verb-friendly languages (e.g., T. Tardif, 1996). The results also suggest that the input language uniquely shapes verb learning such that English-speaking children require grammatical support to learn verbs, whereas Chinese children require pragmatic as well as grammatical support. This research bears on how universally shared cognitive factors and language-specific linguistic factors interact in lexical development.

Understanding how children learn nouns and verbs is crucial for our understanding of lexical development. Nouns typically designate objects that are visually stable across time and are individuated from the environment. In contrast, verbs typically denote actions that are more ephemeral and more difficult to individuate than objects. How do children learn nouns and verbs in the face of these conceptual differences?

This question is deeply related to the issue of whether noun learning is favored over verb learning across languages. Gentner (1982) proposed that nouns are easier to learn than verbs independent of the structural and distributional properties of the native language. However, this proposal has been challenged by results showing that early lexicons may contain equal numbers of nouns and verbs if verbs predominate in the input (e.g., Choi & Gopnik, 1995; Tardif, 1996). In languages including Chinese, Korean, and Japanese, both subjects and objects are often dropped from the sentence, having the effect of making verbs more prevalent in the input. Importantly, verbs appear more frequently than nouns in the maternal input (e.g., Choi & Gopnik, 1995; Kim, McGregor, & Thompson, 2000; Ogura, 2001; Tardif, 1996).

The question of whether Chinese, Korean, and Japanese children learn verbs earlier or at the same time as nouns is far from settled as mixed results have been reported. Some studies suggest that verbs predominate in the early production vocabularies of

---

This research was supported by the Ministry of Education grant-in-aid for Scientific Research awarded to Imai (Grants 15300088 and 18300089), Haryu (Grant 17330139), and Keio Gijuku Academic Development Funds and Keio Gijuku Mori Memorial Research Fund to Imai. We thank the children, teachers, and parents for their participation in this research. We are indebted to Hua Shu for insightful comments on earlier versions of the manuscript as well as invaluable suggestions in constructing nonsense words and instructions in Chinese. We also thank Dedre Gentner and three anonymous reviewers for insightful comments on earlier versions of the manuscript. Hirsh-Pasek and Golinkoff thank their laboratory coordinators, Meredith Jones and Amanda Brandone for their help in carrying out this project and support from National Institute of Child Health and Human Development Grant 5R01HD050199 and by National Science Foundation Grant BCS-0642529.

Correspondence concerning this article should be addressed to Mutsumi Imai, Faculty of Environment and Information Studies, Keio University at Shonan-Fujisawa, 5322 Endo, Fujisawa, Kanagawa, 252-8520, Japan. Electronic mail may be sent to imai@sfc.keio.ac.jp.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7904-0012

Mandarin- and Korean-speaking children (Choi, 2000; Tardif, 1996). Other studies report approximately equal proportions of nouns and verbs (Choi & Gopnik, 1995). Still other studies note that the proportion of nouns is *higher* than verbs in Korean (Au, Dapretto, & Song, 1994; Bornstein & Cote, 2004; Kim et al., 2000) and in Japanese (Ogura, 2001; Yamashita, 1997) children's vocabularies.

Given the results reported in the literature, the issue of noun – verb dominance remains open. Additional investigations of some of the "verb-friendly" languages such as Navaho (Gentner & Boroditsky, 2005) and Tzeltal (Brown, Gentner, & Braun, 2005) as well as microlevel analyses of Chinese children's noun and verb vocabulary through different periods of lexical development are underway (Tardif, 2006). Nevertheless, a different approach is needed to corroborate the research using the checklists or corpus data because data from these methodologies have limitations.

Checklist and Corpus Data: Limitations

Methods of data collection that involve checklists and corpora provide us with rich and important information on the nature of early vocabularies. At the same time, each of these methods raises a number of issues that need to be addressed. First, context matters for corpus data. In studies using production data, the relative proportion of nouns and verbs in children's early vocabulary is taken to be an index of the relative difficulty of noun and verb learning. However, the relative proportion of nouns and verbs varies depending on the context in which the production data are collected even within a single language (Choi, 2000, Ogura, 2001; Tardif, Gelman, & Xu, 1999). For example, Japanese mothers produced more verbs in a toy-play context, whereas more nouns were produced in a picture book reading context (Ogura, 2001). Thus, samples collected in different contexts may yield different results.

Second, what counts as a "word" may differ across languages, and this may influence the number of nouns and verbs appearing on checklists or coded from production data. This problem influences the count for verbs more than for nouns because verbs tend to be more polysemous and crosslinguistically variable in their meanings (e.g., Gentner, 1982). For example, 开 (kāi) is a monosyllable verb in Chinese. Its meaning includes the rough equivalent of "to open," "to start," "to operate/drive," and "to release," among others. Even though 开 has its own entry as a verb in Chinese dictionaries, Chinese dictionaries usually list each word combination as a separate lexical entry (e.g., 开门 (kāimén) "open (a/the) door," 开会 (kāihuì) "have (a/the) meeting," 开车 (kāiche) "to drive"). In contrast, in English, although frequent verbs such as "run" and "take" are also polysemous, English speakers do not consider different senses of these words to be different lexical items. Thus, in constructing a checklist inventory in English, a polysemous verb is usually listed only once as a single lexical item.

Third, the use of a word (as indicated on a checklist or in production) does not guarantee full mastery of the meaning of the word. Children may use a particular word in a particular situation appropriately, but the total number of situations in which they could use the word appropriately may be much more limited than situations in which the word is used by adults (Bowerman, 1980; Dromi, 1987). For example, a child may use the verb "throw" when she sees someone throwing a ball, but she may not fully understand that one can throw any object that can be held with one's hand and one can throw things in many different ways (e.g., with two hands or one hand, underhand, or overhand, etc.; Forbes & Poulin-Dubois, 1997; see also Huttenlocher, Smiley, & Charney, 1983). Thus, although the age of an entry on a checklist or in production has been a popular measure of acquisition of word meaning, it may take a long time for children to acquire an adult-like understanding of the meaning of a word after it is first entered in the child's vocabulary (Theakston, Lieven, Pine, & Rowland, 2002).

Considering these issues, it is difficult to conclusively determine whether nouns or verbs are learned more "easily" based on the checklist and production data alone. Furthermore, characterizing noun and verb learning along the dimension of "ease/difficulty" of learning may not be the best way to deepen our understanding of the *nature* of the early lexicon and of the *mechanisms* that undergird early lexical development.

Here, we investigate the nature of early noun and verb learning from a different perspective. Specifically, we ask whether children from three different language groups are able to map and extend novel nouns and novel verbs, and how conceptual differences between nouns and verbs interact with the linguistic and distributional properties of a target language. Using a laboratory-based, novel word-learning paradigm, we supplement what is known about the process involved in the learning of novel nouns and verbs. Importantly, this is a different (and perhaps complimentary) way to approach these issues as both checklists and production data assess words that are already evident to some extent in children's vocabulary. Studying how readily young children from three language groups map and extend novel nouns and verbs adds a new dimension to our understanding of the mechanisms that support word learning and to how a child's native language interacts with these mechanisms.

Mapping and Extending Novel Nouns and Verbs

Tests of word extension (e.g., knowing that a cup refers to more than the original cup) clarify the construal children give for a particular lexical item. For example, if extension is inappropriately limited, it suggests that children have an overly narrow meaning for the novel word that might be limited to a particular context. Further, names for objects are often generalized on the basis of the sameness of shape (Imai, Gentner, & Uchida, 1994; Landau, Smith, & Jones, 1988). Names of substances (like sand and water), however, must be generalized on the basis of the sameness of the material alone. Temporary configuration of the substance is irrelevant for generalization (Imai & Gentner, 1997; Imai & Mazuka, 2003, 2007; Soja, Carey, & Spelke, 1991). The meaning of a noun—whether it is a name of an object or of a substance—is independent of how the referent entity is acted upon. A ball is a ball whether it is thrown, kicked, rolled, etc.

Verbs are extended to categories of relations. An action can be done with many different objects. We can throw a ball, a Frisbee, a stone, a disk, or almost anything we can lift up with our hands. Thus, in generalizing a verb, children need to align relevant components of action events, compare across different scenes, and focus only on the sameness of the higher order relation between the objects while ignoring the sameness of the objects (Imai, Haryu, & Okada, 2005).

How do children map novel nouns and verbs onto nonlinguistic conceptual categories? Childers and Tomasello (2002) examined whether 2-year-old English-speaking children appreciate both the noun–object link and the verb–action link. They showed 2.5-year-old English-speaking children three different novel objects and taught their noun labels. In another condition, they demonstrated three novel actions performed with three novel objects and taught three novel verbs corresponding to the three actions. Their results indicated that children learned the noun–object link more readily than the verb–action link. Their study, however, did not ask how children extend novel nouns and verbs.

Golinkoff, Hirsh-Pasek, Bailey, and Wenger (1992) and Golinkoff, Jacquet, Hirsh-Pasek, and Nandakumar (1996) studied fast-mapping and extension of novel nouns and verbs. In the first study with nouns (Golinkoff et al., 1992), children were asked to select a novel object amid three familiar objects when it was requested with a novel noun. On the next trial, they were asked to extend the novel noun, selecting from among two familiar exemplars (e.g., a shoe and a block), another novel exemplar, and the original exemplar in another color. At 28 months, children could learn and extend a novel noun in this task. Using cards with Sesame Street characters depicting familiar and novel actions, it took an additional 6 months for young children to learn and extend novel verbs (Golinkoff et al., 1996).

Imai et al. (2005) asked how Japanese children differentially learn and extend nouns and verbs. Imai et al. showed a video of an action event involving a woman, a novel action, and a novel object. Half of the children heard a novel noun, and the remaining children heard a novel verb while they were watching the video. The children were presented with two variants of the standard event: In one, the same woman was doing the same action with a different object; in the other, the same woman was doing a different action with the same object as the standard event. Children were asked to indicate to which scene the word (the newly introduced noun or verb) would apply. Both 3- and 5-year-olds successfully mapped a novel noun to an object, generalizing it to the same object used in a different action. In stark contrast, only 5-year-olds, but not 3-year-olds, showed understanding that verbs are extended on the basis of the sameness of actions and the objects are variables that can be changed across different events (see Kersten & Smith, 2002, for converging results with English-speaking children).

These studies suggest that children more readily map and extend novel nouns than novel verbs. However, none of these studies included a comparison across different language groups. A crucial issue for our understanding of the nature of early lexical development is how word learning interacts with the nature of the language children are learning. It is thus extremely important for us to consider how children's performance in novel noun and verb learning in the same experimental settings interacts with linguistic and distributional properties of their language.

Comparing Word Learning Across Languages: The Present Research

Comparing children learning English, Japanese, and Chinese is theoretically interesting because the three languages differ from one another on dimensions thought to affect the relative difficulty of verb learning. One dimension on which the languages differ is whether they allow argument dropping in the production of the sentence. In Japanese and Chinese, verb arguments are frequently dropped (e.g., the nouns surrounding the verbs are not expressed). In fact, in Japanese, when the speaker thinks that the hearer can infer the arguments from the context, as a default, the arguments are dropped. This means that, in Japanese and Chinese, it is possible that a verb is the only word in the sentence. For example, when you see someone drop a wallet from her purse, in Japanese, you are likely to say "otoshi (drop)-mashi (honorific)-ta (past)-yo (emphatic sentence-final particle)" "(you) dropped (it)." Here, the subject and the object of the sentence are both omitted. In English, this kind of argument dropping virtually never occurs.

The second dimension on which the languages differ is in the presence of morphological inflection as in verbs. Here, Chinese contrasts with not only English but also with Japanese. Although verbs conjugate in both English and Japanese, they do not in Chinese. Thus, nouns and verbs are not *morphologically* distinguished in Chinese (Erbaugh, 1992). In Chinese and Japanese, where verb arguments are often dropped, the verb alone can constitute a sentence in the utterance. In Japanese, even when this occurs, verbs can be identified by inflectional morphology. That is, when a verb is produced without the arguments, as in "Mite (Look), X-teiru (X-ing)," one can tell by the presence of the "teiru" ending that the word X is a verb. However, in Chinese, when a word is produced on its own (and this happens occasionally in conversational discourse), it is difficult to tell whether it is a noun or a verb (Li, Jin, & Tan, 2004). A novel word can be identified as a verb *only when* it is embedded in the argument structure of the prior discourse (see Li, Bates, & MacWinney, 1993, for discussion of how Chinese-speaking adults determine grammatical classes in sentence processing).

This research extends Imai et al.'s (2005) study to examine two questions. First, we probe whether the noun advantage found in Imai et al.'s (2005) study holds across children learning very different languages. Second, we ask whether children speaking languages that have been noted as verb friendly such as Japanese and Chinese more readily learn and extend novel verbs than children speaking English, reputed to be a "noun-friendly" language. Some researchers have claimed that Chinese and Japanese are more verb-friendly languages than English, mainly because of their grammatical as well as lexical distributional properties (Choi & Gopnik, 1995; Ogura, 2001; Tardif, 1996; cf. Gentner, 1982). With argument dropping, verbs should be relatively more frequent in Chinese and Japanese than in English. Additionally, the morphological simplicity of Chinese verbs might give an advantage to Chinese verb learners, as children do not need to learn various inflectional forms of the same verbs (Tardif, 1996).

From a different theoretical perspective such as syntactic bootstrapping hypothesis, the opposite prediction can be made. The syntactic bootstrapping hypothesis asserts that children rely on the syntactic structures in which verbs appear to infer their meanings (e.g., Fisher, 1996; Gleitman, 1990; Naigles, 1990). Granting that Chinese does not have much in the way of morphology, Lee and Naigles (2008) argue that syntactic bootstrapping is possible in Mandarin Chinese because children hear verbs in multiple sentence frames and can rely on regularities in how verbs are used syntactically. For example, certain verb classes are more likely to occur with locative or prepositional phrases, and transitive verbs are reliably signaled by the use of postverbal noun phrases—when they appear. In natural discourse, however, children may not hear a new verb used in different phrasal frames at that very occasion; it is more likely that children hear a verb but in a single frame—as in the present study—at a time. Here, Chinese children may be at a serious disadvantage relative to English-reared children who can exploit the argument structure that is usually present in English for learning and extending a novel verb.

According to the syntactic bootstrapping hypothesis, which emphasizes the role of argument structure for the inference of verb meaning, Chinese and Japanese children might be at an equal disadvantage for *novel* verb learning. On the other hand, if morphological information distinguishing verbs from nouns also matters, the disadvantage for Chinese children may be more serious than for Japanese children.

In the present research, to examine how the richness of morphosyntactic cues affects novel verb learning, we examined variability in argument structure both within and across languages. In English and Japanese, even when the information about argument structure was not provided, the word's form class could be inferred from the morphological information. Chinese thus presents an interesting test case for the use of argument structure and for how the absence of argument structure influences object versus action word mapping because the words without morphology are so ambiguous with respect to form class. In languages that use distinct morphology for nouns and verbs, once children learn the relation between word forms and word meanings (e.g., Golinkoff, Hirsh-Pasek, & Schweisguth, 2001), it is not possible to test whether they are more willing to map a new word to an object or to an action without being influenced by the form in which the word is presented. In Chinese, this is perfectly possible.

Within language, in English and Japanese in particular, we set up two conditions where verbs are introduced with or without arguments. In previous studies with Japanese children (Imai et al., 2005), novel verbs were presented without arguments (e.g., "Mite, X-teiru! "Look!, X-ing!"). The verb was offered in the present progressive tense, which clearly indicated that the word was a verb—at least to adults. However, children might profit from explicit specification of the arguments of the verb even when they can be inferred from contextual cues.

In summary, this research should illuminate the question of what is universal and what is language specific in young children's noun and verb learning, and how children' use of argument structure interacts with linguistic properties of the input language. Six studies were conducted across three languages.

Study 1: Japanese Children's Novel Noun and Verb Learning

Study 1 examined how monolingual 3- and 5-year-old Japanese children map a novel noun and a verb and how they extend it to a new situation, presenting a word in one of the three conditions: novel noun, bare verb, and verb with arguments. Study 1 was in part a replication of Study 1 in Imai et al. (2005), with additional specification of verb arguments. Of particular interest here is whether Japanese children still learn novel nouns more readily than novel verbs when the verbs were presented in full argument structure.

Method

Participants. Forty-one 3-year-old ($M = 3.6$, range $= 3.0 - 4.1$) and forty-two 5-year-old ($M = 5.5$, range $= 5.0 - 6.0$) monolingual, normally developing Japanese children participated in the study. Six additional 3-year-olds were tested but eliminated from the final sample for a side bias (2), for failing to complete the study (2), and for being bilingual (2). (Children were judged to be "side biased" when they selected the video on the same side, either left or right, throughout the six trials.) Within each age group, children were randomly assigned to one of the three conditions: noun, verb with arguments, and bare verb conditions. The proportion of boys and girls was approximately equal within each age group, and the age distribution (the mean age and range) was the same across the three conditions. Children lived in a suburban city in the Greater Tokyo area and were mostly from middle-class families.

Stimulus materials. Six sets of video action events that were identical to those used in Study 1 of Imai et al. (2005) served as stimulus materials. Each set consisted of a standard event and two test events, each lasting approximately 10 s. In the *standard* event, a young woman did a novel repetitive action with a novel object (e.g., a woman holds a brown plastic drainpipe in her right hand and pushes it outward with a punching motion). The two *test* events were variants of the standard event. In one, the same person did the same action, but the object was replaced with an object that was distinctively different from the standard in all perceptual dimensions, particularly in shape (action-same-object-change, henceforth AS). In the other test event, the theme object was the same but the action was distinctively different from the action in the standard event (object-same-action-change, henceforth OS). In both test videos, the actor was the same as the one in the standard picture. We made the action as perceptually stable and salient as possible to assure that young children would be able to decompose the action event into agent, action, and object components. (See Table 1 for the list of the actions and objects used in the six sets.)

Procedure. Participants were tested individually at the preschool they attended by a female native speaker. The experimental procedure was the same as that in Imai et al. (2005) except that the stimulus presentation was controlled by PowerPoint.

As in Imai et al. (2005), children received four warm-up trials to make sure that they could indicate what they thought to be the correct answer by pointing to one of the two test video clips. Care was taken so that the practice trials would not lead to any attentional bias toward the object or action in the event during the test trials. In the first two practice trials, they were shown two familiar objects (e.g., a cat vs. a dog) and were asked to identify the object the experimenter referred to "e.g., a dog." In the next two practice trials, they were shown a person doing two familiar actions. In one trial, for example, a woman was jumping in one video and clapping hands in the other, and children were asked, "Where is she jumping?" The last warm-up trial first presented an actor doing a novel action with a novel object, followed by two test events shown simultaneously and side by side. One was exactly the same as the standard; the other was totally different both in the action and in the object, although the actor was the same. As in the test trials, a novel word (either a noun or a verb,

Table 1
Description of the Actions and Objects Used in Studies 1–6

	Standard	Action for the object same test	Object for the action same test
Set 1	A woman is rolling the object (a football-shaped orange Frisbee with fins) between the palms of her hands.	A woman is lightly tossing and catching the object with both hands.	A blue candleholder in the shape of a cup with a long stem.
Set 2	A woman is holding the object (a long blue plastic arch-shaped object) behind her back, pulling it up and down.	A woman is holding the object in front with both hands, twisting her torso from side to side.	A long narrow meshed black band made of rubber.
Set 3	A woman is holding the object (a brown plastic drainpipe) in her right hand and pushing it outward with a punching motion.	A woman is holding the object in her right hand and tapping it against her left shoulder.	A round metallic timer.
Set 4	A woman is holding the object (a blue rubber dog toy consisting of two large loops), both hands at chest level, twisting the object so that it bends in the middle.	A woman is holding the object in her right hand and slightly moving her hand so that the object flops up and down.	An octagon-shaped rubber Frisbee ring with blue/black stripes.
Set 5	A woman is holding the object (a green metal spherical trellis on a long, thin stem) in her right hand, pushing it out with a stabbing motion.	A woman is holding the object in her right hand and tapping it against her right knee, which she is raising at the same time as she is lowering the object.	A narrow white cylindrical plastic trellis.
Set 6	A woman is tapping the object (a donut shaped plastic toy for cats) against her thigh.	A woman is holding the object with the index fingers of each hand stuck into the hole in the center and rolling the object around her fingers.	A large gray plastic curved pipe.

depending on the condition the child was assigned to) was introduced. The child was asked to which movie the novel word applied.

After the warm-up trials, children received six test trials. The standard event was presented for 30 s during which a novel word was presented. Table 2 gives the instructions for each language. The target word was introduced approximately 2 s after the standard movie began to make sure that the word was said while the actor was doing the action. The word was repeated (in the appropriate frame for the condition) three times for each scene (see the Appendix for a list of the nonsense words).

Two test events were then offered simultaneously and side by side on the same monitor. The participant was asked to which of the two events the target word should be extended. Presentation order of the six sets was randomized, and the location of the AS and OS test events was counterbalanced across the six sets within each participant.

Conditions and instructions. Novel words were presented either as a noun or as a verb. In the *noun* condition, children heard, "Look, this is (a) X!" In Japanese, no singular–plural distinction is made. Thus, in the actual Japanese instruction, children heard, "Kore (this) wa (topic) X desu (IS)," where there is no number marking (see Imai & Haryu, 2001). Verbs were presented in two different forms: one with full arguments (*verb with arguments* condition) and the other with no arguments (*bare verb* condition). In the verb with arguments condition, the word "oneesan ('girl')" is used for the subject, and "nanika ('something')" was used to refer to the novel object ("Oneesan ga nanika o X-teiru" "The girl is X-ing something"). In the *bare verb* condition, the subject and the object were not specified in the sentence, but a novel word could be identified as a verb by the morphological affix "-teiru."

Results

The proportion of the AS responses is shown in Table 3. Children's performance in the noun and bare verb conditions replicated that in the noun and verb

Table 2

Conditions and Instructions for Studies 1–6

Language	Condition	Instruction during verb presentation	Instruction for test
English	Noun (Study 2)	"Look! This is a X!"	"Where is the X? Can you point to the X?"
	Bare verb (Study 2)	"Look! X-ing"	"Where is X-ing?"
	Verb with arguments (Studies 2, 6)	"Look! She is X-ing it"	"Where is she X-ing it?"
Japanese	Noun (Study 1)	"Mite (look)! X-ga (Nominal particle) aru (exist)" "Look! There is (a) X"	"X-ga aru (exist)-no (Nominal particle)-wa (Topic particle) docchi (which movie)?" "In which (movie) is there (a) X"
	Bare verb (Studies 1 and 6)	"Mite (look)! X-teiru (X-progressive)" "Look, X-ing"	X-teiru-no(Genetive particle)-wa (Topic particle) docchi (which movie)?" "In which (movie) is (she) X-ing?"
	Verb with arguments (Study 1)	"Mite (look)! Oneesan (girl) ga (Nominal particle) nanika (something)- wo (Accusative particle) X-teiru (X-progressive)" "Look, she is X-ing something"	"Oneesan (girl) ga (Nominal particle) nanika (something)- wo (Accusative particle) X-teiru (X-progressive) no (Genetive)-wa (Topic) docchi (which movie)?" "In which (movie) is she X-ing something?"
Chinese	Noun (Studies 3, 3A, 5)	"Ni (you) kan (look)! Nali (there) you (exist) ge[a] (classifier) X" "Look! There is (a) X"	"Na (which) zhang (the quantifier) tu (picture) li (within) you (exist) ge (classifier) X?" "In which picture is there (a) X?"
	Verb with arguments: "zai" only (Studies 3, 3A, 4)	"Ni (you) kan (look)! Ayi (girl) zai (progressive) X yi (one) ge (classifier) dongxi (thing) ne (mode marking particle)" "Look (a) girl is X-ing one thing"	"Na (which) zhang (classifier) tu (picture) li (within) ayi (aunt) zai (progressive) X yi-(one) ge (classifier) dongxi (thing)?" "In which picture is she X-ing something?"
	Verb with arguments: three sentences with different auxiliaries (Studies 4, 5)	"Ni (you) kan (look)! Ayi (girl) zai (progressive) X yi (one) ge (classifier) dongxi (thing) ne (mode marking particle)" "Look (a) girl is X-ing one thing" "Ni (you) kan (look)! Ayi (girl) zhengzai (progressive) X yi (one) ge (classifier) dongxi (thing) ne (mode marking particle)""Look (a) girl is X-ing one thing" "Ni (you) kan (look)! Ayi (girl) yizhizai (progressive) X yi (one) ge (classifier) dongxi (thing) ne (mode marking particle)""Look (a) girl is always X-ing one thing"	"Na (which) zhang (classifier) tu (picture) li (within) ayi (aunt) zai (progressive) X yi-(one) ge (classifier) dongxi (thing)?" "Ayi (aunt) zai (progressive) X yi-(one) ge (classifier) dongxi (thing) de (progressive) tu (picture) shi (is) na (which) yi-(one) ge(classifier)" "In which picture is she X-ing something?"
	Bare word (Studies 3, 3A, 5)	"Ni (you) kan (look)! X!" "Look, X!"	"Na (which) ge (classifier) tu (picture) zhi (indicate) de (Nominal particle, auxiliary word) shi (is) X? " "Which picture shows X?"

[a] Note that in the Japanese instructions, the noun (either the novel noun in the noun condition or "nanika" (something) in the verb with arguments condition) was not accompanied with a classifier, whereas in Chinese, these nouns were accompanied by the generic classifier "ge." In Japanese, the number of the object is mentioned only when it is pragmatically important in the discourse. Thus, it was very unnatural to use a classifier in this situation. In contrast, in Chinese, "yi-CL-X (noun)" phrase roughly corresponds to English "a X." In our instructions, it was more natural to say "yi (one)-ge (CL)-X" than to say the bare noun without a classifier. Also, the classifier made it clear that X was a noun in the noun condition. However, the presence of the classifier does not highlight the object in the sentence in Chinese.

conditions in Study 1 of Imai et al. (2005): Both the Japanese 3- and 5-year-olds were able to map novel nouns to the OS scene, 93% and 100%, respectively. In the bare verb condition, 5-year-olds selected the AS response significantly above chance (77.4%), $t(13) = 2.43$, $p < .05$, two-tailed, but 3-year-olds were at chance level (38.1%), $t(13) = -1.072$, $p > .05$. In the verb with arguments condition, both 3-year-olds (39.3%) and 5-year-olds (69%) were at chance levels, $t(12) = -0.90$ and $t(13) = 1.56$, both $ps > .05$, respectively.

To examine whether the pattern of the results of the group analysis holds for the distribution of individuals, we counted the number of the children who selected the AS test four or more times as AS choosers (see Table 4). The results of the individual analysis converged with the group analysis; the number of AS choosers was significantly below chance by the binomial criterion in the noun condition for both ages, $p < .01$. For the 3-year-olds, the number of the AS choosers did not exceed chance either in the bare verb or in the verb with arguments condition. For the 5-year-olds, it exceeded chance in the bare verb condition in 5-year-olds, $p < .05$, but not in the verb with arguments condition, $p > .1$. In this and the following studies (Studies 2–6), the proportion of the AS choosers with the results of the binomial tests against chance is reported in Table 4.

A 2 (age) × 3 (condition) analysis of variance (ANOVA) was conducted on the proportion of the AS responses to examine the contribution of the two factors of age and the syntactic frame in which the novel word was presented (condition). A main effect for age, $F(1, 77) = 6.79$, $p < .05$, and condition, $F(2, 77) = 19.27$, $p < .01$ emerged, but the main effects were qualified by a significant Age × Condition interaction, $F(2, 77) = 3.16$, $p < .05$. To decompose the interaction, three separate 2 (age) × 2 (condition) ANOVAs were conducted. The Age × Condition interaction emerged in the analysis directly contrasting the noun and bare verb condition, $F(1, 51) = 8.02$, $p < .01$, as well as that contrasting the noun and verb with argument condition, $F(1, 51) = 4.37$, $p < .05$, but not in the one contrasting bare verb and verb with arguments conditions, $p > .05$. In the last analysis, the main effect of condition was not significant, either, $p > .05$.

Discussion

Replicating the results of Imai et al. (2005), both 3- and 5-year-old Japanese children mapped a novel noun to a novel object in an event and applied the word to the scene in which the same object appeared in a different action. Another way in which this

Table 3

Proportion of Action-Same-Object-Change Responses in Study 1 (Japanese), Study 2 (English), Study 3 (Chinese), and Study 6 (Japanese- and English-Reared Children Tested on Stimuli in Which the Object-Holding Segment Was Removed)

			Condition			
Language	Study	Age	Noun	Bare verb	Verb with argument	Bare word
Japanese	Study 1	3-year-olds	0.07**,[a] (0.11)	0.38 (0.41)	0.39 (0.44)	—
		5-year-olds	0.00** (0.00)	0.77* (0.42)	0.69 (0.45)	—
	Study 6	3-year-olds	—	0.52 (0.33)	—	—
English	Study 2	3-year-olds	0.14** (0.15)	0.49 (0.39)	0.42 (0.35)	—
		5-year-olds	0.09** (0.12)	0.56 (0.42)	0.70* (0.33)	—
	Study 6	3-year-olds	—	—	0.53 (0.27)	—
Chinese	Study 3	3-year-olds	0.15** (0.23)	—	0.08** (0.15)	0.15** (0.23)
		5-year-olds	0.06** (0.19)	—	0.20* (0.38)	0.22* (0.33)
	Study 3A	Adults	0.29* (0.37)	—	1.00** (0.00)	0.73* (0.24)

[a] If the number is smaller than .5 and has * or **, the object-same response was significantly above chance.
$*p < .05$. $**p < .01$.

Table 4

Proportion of the Action-Same Choosers in Study 1 (Japanese), Study 2 (English), Study 3 (Chinese), and Study 6 (Japanese and English, Tested on Stimuli in Which the Object-Holding Segment Was Removed)

			Condition			
Language	Study	Age	Noun	Bare verb	Full argument verb	Bare word
Japanese	Study 1	3-year-olds	0**[a] (0/14[b])	0.49 (5/13)	0.39 (6/14)	—
		5-year-olds	0** (0/14)	0.79* (11/14)	0.64 (9/14)	
	Study 6 (Object-holding segment removed)	3-year-olds	—	0.39 (6/14)	—	—
English	Study 2	3-year-olds	0** (0/18)	0.35 (6/17)	0.32 (6/19)	—
		5-year-olds	0** (0/20)	0.5 (9/18)	0.76* (16/21)	—
	Study 6 (Object-holding segment removed)	3-year-olds			0.38 (6/16)	
Chinese	Study 3	3-year-olds	0.08** (1/12)	—	0** (0/12)	0.17** (2/12)
		5-year-olds	0.08** (1/12)	—	0.17** (2/12)	0.17** (2/12)
	Study 3A	Adults	0.17** (2/12)	—	1.0** (12/12)	0.6 (12/18)

[a] If the number is smaller than .5 and has * or **, the object-same response was significantly above chance.
[b] The numerator represents the number of the action-same choosers; the denominator is the *n* in the condition.
$^{*}p < .05$. $^{**}p < .01$.

research replicated and extended the previous results was in demonstrating that 3-year-olds could not map and extend novel verbs to the same action in the face of an object change, whether the verb was presented with or without the specification of the arguments. It was somewhat unexpected that Japanese 5-year-olds failed to extend a novel verb to the same action scene reliably above chance when the verb was presented with the arguments but succeeded when the verb was presented without the arguments. Given that the difference between the two verb conditions (verb with arguments and bare verb) did not reach the level of statistical significance, we could not confidently conclude that Japanese children learn verbs better when unnecessary arguments are dropped. However, what is clear is that Japanese children did not gain from the specification of argument structure *when the arguments could be easily inferred from the context*. The action events used in our research involved only three elements: an actor, an action, and an object. From the Japanese point of view, it was obvious that the subject was the actor and the theme object was the novel object; hence, it was more natural that the arguments be dropped. Japanese children, in fact, could have been distracted by hearing this unnecessary information.

In Study 2, we test English-speaking 3- and 5-year-olds in the same three conditions (noun, bare verb, and verb with arguments) to compare their performance. Even though English has been characterized as a noun-friendly language, it is possible that English-speaking children learn verbs more readily than, or as readily as, Japanese children, as English offers an important structural cue to novel verb meaning, namely, argument structure. Furthermore, English-speaking children should learn and extend novel verbs more readily when the verb is presented with the arguments but not when it is presented in the bare form, as the former is much more "natural" than the latter for English speakers.

Study 2: English-Speaking Children's Novel Noun and Verb Learning

Method

Participants. Fifty-four 3-year-old ($M = 3.5$, range = 3.0–3.11) and fifty-nine 5-year-old ($M = 5.2$, range = 4.5–5.11) monolingual, normally developing English-speaking children in the suburban Greater Philadelphia area, mostly Caucasian from middle-class to upper-middle-class families, participated. An additional 21 children were tested but were not included for experimenter error (5), parental interference (4), bilingual knowledge (3), and failure to cooperate (9). The children were tested individually by a native speaker of English in two Northeastern University laboratories and were randomly assigned to the three conditions: noun, bare verb, and verb with arguments. Gender distribution and age distribution (the mean age and the age range) were approximately equal within each age/condition group.

Stimuli and procedure. The stimuli and the procedure were the same as those used for the Japanese children in

Study 1. In the noun condition, a novel word was presented as a noun ("Look, this is a X!"). In the bare verb condition, novel verbs were presented with a present progressive /ing/ ending, but the subject and the object of the sentence were dropped ("Look, X-ing!"). In the verb with arguments condition, the pronoun "she" served as the subject and "it" as the object of the sentence (e.g., "Look, she is X-ing it"). For the test, the experimenter asked the children "Where is the X?" "Where is X-ing?" "Where is she X-ing it?," respectively.

Results and Discussion

Table 3 shows the mean proportions of the AS responses by 3- and 5-year-old *English-speaking* children in all three conditions. Both the 3- and the 5-year-olds in the noun condition successfully mapped novel nouns to novel objects, reliably selecting the OS event, 85.9%, $t(17) = 9.88$, $p < .001$, for the 3-year-olds, and 89.8%, $t(19) = 14.5$, $p < .001$, for the 5-year-olds. In extending novel verbs, the 5-year-olds in the verb with arguments condition successfully selected AS events 70.1% of the time, $t(20) = 2.8$, $p < .02$. However, the 5-year-olds in the bare verb condition performed at chance, 55.6%, $t(17) = .56$, $p > .1$. The performance of the 3-year-olds did not differ from chance, whether they were in the verb with arguments condition or in the bare verb condition, 42.3%, $t(18) = 1.18$, $p > .1$ and 48.5%, $t(16) = .35$, $p > .1$, respectively. Individual analyses were also conducted, and the results converged with those of the group analysis (see Table 4).

We then examined whether English-speaking children's AS responding differed across the three conditions, and whether the effect interacted with age. The effect of condition was significant, $F(2, 107) = 23.13$, $p < .001$, as well as the Age × Condition interaction, $F(2, 95) = 3.83$, $p < .05$. As in Study 1, we conducted three 2 (age) × 2 (condition) ANOVAs to directly examine the source of the significant Age × Condition interaction. In the ANOVA contrasting the two verb conditions, the Age × Condition interaction was significant, $F(1, 71) = 4.01$, $p < .05$. In the analysis contrasting the noun and the verb with arguments condition, the interaction was also highly significant, $F(1, 71) = 7.34$, $p < .01$. However, in the analysis contrasting the noun and the bare verb condition, the Age × Condition interaction effect did not reach the level of significance.

Thus, the results of ANOVA analyses were consistent with the overall pattern suggested by the means for each age and condition. English-speaking 3- and 5-year-olds could fast-map a novel noun to an object, just as the Japanese children did in Study 1. In contrast, 3-year-old English-speaking children were not willing to extend a novel verb to an event in which the same action was performed with a different object, paralleling the findings with the Japanese 3-year-olds. The English-speaking 5-year-olds could extend a novel verb properly but only when verbs were presented with the full argument structure and not when they were presented in the bare frame. In other words, 5-year-old children from both language groups would extend a novel verb only when the verb was presented in the structural form that was most typical in their language. These results suggest that language specificity matters for the acquisition of verbs. In learning and extending verbs, children's responses are affected by the type of verb input they hear.

Study 3: Chinese Children's Novel Noun and Verb Learning

We now consider how Chinese children, learning a language that does not differentiate between nouns and verbs through the use of inflectional morphology, perform in this task. Of particular interest is whether Chinese children are able to map and extend novel nouns as well as Japanese- and English-speaking children and whether they do so at a rate better than that achieved with novel verbs, even though they tend to hear fewer nouns in the input than English-reared children. If Chinese children do manifest the noun advantage, this will provide strong support for the view that universal cognitive factors are more prominent than language-specific factors in influencing the ease and willingness with which children learn novel words. The Chinese data will also illuminate the role of syntactic cues in novel verb learning. Chinese children are offered syntactic cues less systematically than children learning either Japanese or English, both in terms of the absence of morphological marking on verbs and the occasional absence of argument structure.

Because nouns and verbs are not morphologically distinguished, the condition equivalent to the bare verb condition in Studies 1 (Japanese) and 2 (English), in which both the subject and the object of the sentence are omitted, is not achievable; a novel word cannot be presented as a verb in its bare form. Unless embedded in argument structure, it is unclear how to assign a word to a form class (see the Discussion section). We thus did not have the bare verb condition for the Chinese group. Instead, we included a *bare word* condition, in addition to the noun and the verb with arguments condition because the bare word condition in Chinese presents an ideal case to address a different but related issue. That is, Chinese allows us to ask whether young children have a predisposition to name a novel object rather than a novel action (Gentner, 1982; Golinkoff et al., 1992; Markman, 1989). If Chinese children map the novel word to the object rather than the action, it suggests that children wish to name objects in the environment independent of the frequency of verbs in the linguistic input. On the other hand, if Chinese children map the word to the action, it suggests that the type of the concept that young children are most willing to name depends on the input language.

Method

Participants. Thirty-six 3-year-old ($M = 3.5$) and thirty-six 5-year-old ($M = 5.6$) monolingual, normally developing Mandarin Chinese–speaking children living in Beijing, China, participated. Three other 3-year-olds were tested but were not included in the analysis because they were unwilling to cooperate. The participants attended a university preschool in Beijing. Their parents were mostly from middle and upper-middle class, which is comparable to the demographic background of the Japanese and American children tested in Studies 1 and 2. The children were randomly assigned to one of the three conditions: noun, verb with arguments, bare word. Gender distribution and age distribution (the mean age and range) were approximately equal within each age/condition group. They were tested individually by two female native speakers of Mandarin Chinese in a quiet room.

Stimuli and procedure. The stimuli and the overall procedure (except for the conditions and instructions) were the same as those used for the Japanese children and English children in Studies 1 and 2.

Conditions and instructions. Chinese children were randomly assigned to one of the three conditions: noun, verb with arguments, and bare word. The instructions are given in Table 2. In the noun and the verb with arguments conditions, special care was taken so that there was no ambiguity over whether the target word was a noun or a verb, respectively. In the verb with arguments condition, an auxiliary verb "zai," which marks the imperfective aspect and is usually used in expressing an ongoing action, accompanied the verb along with the subject "ayi ("the auntie [woman older than the hearer]") and the theme object "yi- (one) ge (generic classifier) dongxi (thing)." We confirmed with several preschool teachers that 3-year-old children would easily understand this construction.

Because most Chinese words consist of two syllables, we constructed two-syllable nonsense words created from combinations of existing morphemes (see the Appendix for the list of the words). When spoken, the nonsense words were not similar to existing words.

Results

As shown in Table 3, Chinese children at both ages selected AS events significantly below chance in all three conditions, all *ps* $< .05$ (see also Table 4 for the results of the individual analysis.) The rate of AS responding did not differ across the two age groups or across the three conditions. The interaction between the two factors was not significant. The results indicated that Chinese-speaking children, both 3- and 5-year-olds, have a strong tendency to fast-map a novel word to a novel object, whether the word was presented as a noun or a verb, or as a bare word, whose grammatical form class was not revealed.

Study 3A: Chinese Adults

Given the unexpected results from Chinese children, we tested Chinese-speaking adults to see how they would perform. If the children's strong bias to map a novel word to a novel object even in the verb with arguments condition was due to some unexpected factors in the instruction, Chinese-speaking adults may also show difficulty in mapping the word to the action in this condition.

Method

Participants. Forty-two undergraduate students of Peking University, all whom were native speakers of Mandarin Chinese, were individually tested. Eighteen were tested in the bare word condition and 12 each in the noun and verb with arguments conditions.

Stimuli and procedure. The stimuli and the procedure were the same as in Study 3 except that the warm-up trials were omitted. The instructions and the novel words were the same as those used for Study 3 with preschool children.

Results

Chinese adults mapped a novel noun to a novel object (70.8%), $t(11) = 2.55$, $p < .05$ (see Table 3). In sharp contrast to the preschool age children, adults in the verb with arguments condition selected the AS test 100% of the time. Also different from the children, adult Chinese speakers in the bare word condition made AS response 73% of the time, mapping the class-ambiguous word to the action rather than to the object, $t(17) = 4.03$, $p < .01$.

Discussion of Study 3 and Study 3A

The novel word presented in the verb with arguments condition was construed as a verb by Chinese adults 100% of the time. It is also interesting that the Chinese adults preferred to interpret the word as referring to the action rather than the object when the word's form class was ambiguous in the bare word condition. With the current data, it is not clear whether this apparent action-naming bias in Chinese adults is a consequence of a genuine preference for action-naming over object-naming: Chinese adults may have thought that the verb interpretation was more likely than the noun interpretation because they could come up with a good-enough name for the object (e.g., a superordinate category name such as "tool" or "toy"), but not for the action, for example. In any case, interpreting a novel word as a verb, given our stimuli and instructions in the verb with arguments condition is a perfectly reasonable course of action for adults. Even in the absence of syntactic information for verbhood (morphology or argument structure) in the bare word condition, adults did not make an object-name interpretation. In this light, Chinese children's strong object-naming bias is striking.

Taken together, it seems that noun learning is privileged independent of the properties of the language children are learning (e.g., Gentner, 1982). These results indicate that the linguistic properties of Chinese—argument dropping and morphological simplicity—that have been said to foster verb learning (Tardif, 1996, 2006; cf. Gentner, 1982) do not necessarily lead to more readiness to map and extend novel verbs in young Chinese children when compared to English- and Japanese-speaking children.

Why did Chinese children fail to learn verbs in our task, although they have many verbs in their vocabulary? What aspects of our task made it so difficult for Chinese children to learn novel verbs? We had no concern that the sentence construction used was a problem. When we embedded familiar words in the same construction, children had no problem understanding the sentence. One possible concern is that the instruction in the full argument structure condition may not have been "natural" to Chinese children, given that Chinese is also an argument-dropping language like Japanese. We originally considered a condition in which the target verb was presented with the aspect-marking auxiliary "zai" but without the arguments, which parallels to the bare verb condition in Studies 1 and 2. However, Chinese native speakers all objected to this idea, saying that presenting a *novel* word as a *verb* only with the aspectual auxiliary without both the subject and the object dropped in this situation was too "unnatural." Dropping both arguments occurs occasionally when presenting a familiar verb. Introducing a novel word without both arguments and at the same time making clear that it was a *verb*, however, would be difficult. This is probably due to the fact that unlike Japanese, Chinese lacks verb-inherent morphology. We had run a condition in which the object of the sentence (yi-ge-dongxi "a thing") was dropped but the subject was explicitly said, with 3- and 5-year-olds (12 in each age group). Performance was virtually identical to that in the verb with arguments condition. Given this, we think it unlikely that the poor performance of the Chinese children in the verb with arguments condition could be because unnecessary arguments were provided with the verb in the instruction sentence.

One thing we did note, however, was that the phonological structure of the words may have biased children against a verb interpretation. We constructed two-syllable nonsense words because bisyllabic words are most frequent in Chinese and are neutral with respect to whether they are nouns or verbs. However, verbs denoting simple and basic motions such as "run" and "jump" tend to be monosyllabic. Perhaps, then, the use of bisyllabic words biased children against the verb interpretation.

To test this possibility, in Study 4, we replicated the verb with arguments condition using monosyllabic nonsense words. Given the extreme difference between the Chinese preschoolers and the adults in the verb with arguments condition, we also tested 6- and 8-year-olds.

Study 4: Chinese Children's Novel Verb Learning With Monosyllabic Words

Method

Participants. Twelve 3-year-old (*M* = 3.6, range = 3.1 – 3.11), twelve 5-year-old (*M* = 5.7, range = 5.2 – 5.11), fifteen 6-year-old (*M* = 6.9, range = 6.5 – 7.2), and fifteen 8-year-old (*M* = 8.8, range = 8.3 – 9.3) Chinese middle-class children who were normally developing, monolingual native speakers of Mandarin Chinese, participated. Three- and 5-year-olds were from the same preschool in Beijing as the children in Study 3 but had not participated in it; 6- and 8-year-olds were from an elementary school in Beijing. The children were tested by a native speaker of Mandarin Chinese.

Stimuli and procedure. The stimuli and the procedure were the same as that used in the verb with arguments condition in earlier studies except that the nonsense words were monosyllabic. Nonsense words were made of existing phonemes in Chinese so that they sounded like Chinese words, but a nonexisting tone for the phoneme was applied (e.g., jin2, see Appendix for a list of the words used).

Results and Discussion

As shown in Table 5, the use of monosyllabic nonsense words did not drastically change the preschoolers' performance in novel verb learning: Three-year-olds still made AS responses only 25% of the time, $t(11) = 2.63$, $p < .05$, and 5-year-olds did so 33% of the time, $t(11) = 1.17$, $p > .05$. Six-year-olds' performance still did not exceed chance level (52%), $t(11) = 0.19$, $p > .1$. At 8 years of age, Chinese children finally made AS choices significantly above chance (72%), $t(14) = 2.17$, $p < .05$. A significant age difference was found, $F(3, 50) = 3.35$, $p < .05$, and post hoc pairwise comparisons (Tukey) detected a significant difference between 3-year-olds and 8-year-olds, $p = .035$, and a marginally significant difference between 5-year-olds and 8-year-olds, $p = .095$. The results of the individual analysis (i.e., the distribution of children classified as the AS choosers) are provided in Table 6.

Study 4A: Chinese Preschoolers' Verb Learning With Monosyllabic Words Presented in Three Sentence Frames

The results from Chinese children so far indicate that they have a strong object-naming bias. This bias may contribute to making novel verb mapping and extension less readily available in Chinese relative to Japanese and English.

However, it is hard to believe that Chinese 5-year-olds, if not 3-year-olds, cannot succeed in our task. Even granted that they are biased toward object-naming, there must be conditions in which they can correctly interpret the novel word as a verb and extend it to the same action. Perhaps, even though our instructions worked with Chinese-speaking adults, Chinese children need more linguistic scaffolding. In Study 4A, we tested whether Chinese 3- and 5-year-olds were able to map and extend novel verbs to the same action when additional linguistic cues were provided.

Method

Participants. Fifteen 3-year-old ($M = 3.6$, range = 3.1–3.10) and fifteen 5-year-old ($M = 5.3$, range = 5.1–5.8) monolingual Mandarin Chinese–speaking children who had not been tested in the earlier studies participated.

Linguistic cues. In the verb with arguments condition in the main study, the novel word could be unambiguously identified as a verb by the structure of the sentence, in particular, by the word order and the presence of the aspect marker "zai." However, this linguistic cue may not have been strong enough to indicate the word was a verb, given that the form of the word could not be distinguished from a noun by morphology. We thus presented the verb in three different sentences using three different auxiliaries, namely, "zai," "zhengzai," and "yizhizai," all which mark the progressive aspect, to provide even clearer and stronger clues that the novel word was a verb (see Table 2 for full instructions).

Stimuli and procedure. The stimuli were the same as those in the main study, except for the linguistic stimuli.

Results and Discussion

As shown in Table 5, even when the novel verb was presented in three different sentence

Table 5
Proportion of Action-Same Responses in Chinese Children in Studies 4–5

		Condition			
Stimuli	Age	Full argument verb one syllable word "zai" only	Full argument verb one-syllable word three sentence frames	Bare word	Noun
Original	3-year-olds	0.25*[a] (0.33), Study 4	0.34 (0.299), Study 4A	—	—
	5-year-olds	0.33 (0.49)	0.39 (0.43)	—	—
	6-year-olds	0.52 (0.45)	—	—	—
	8-year-olds	0.72* (0.39)	—	—	—
Object-holding segment removed	3-year-olds	—	0.41 (0.30), Study 5	0.28* (0.12), Study 5	0.33* (0.14), Study 5
	5-year-olds	—	0.88** (0.15)	0.37 (0.41)	0.27* (0.37)

[a] If the number is smaller than .5 and has * or **, the object-same response was significantly above chance.
*$p < .05$. **$p < .01$.

Table 6

Proportion of the Action-Same Choosers in Studies 4–5 With Chinese Children

		Condition			
Stimuli	Age	Full argument verb one syllable word "zai" only	Full argument verb one syllable word 3 sentence frames	Bare word	Noun
Original	3-year-olds	0.25*[,a] (3/12[b]), Study 4	0.27* (4/15), Study 4A	—	—
	5-year-olds	0.33 (4/12)	0.33 (5/15)	—	—
	6-year-olds	0.53 (8/15)	—	—	—
	8-year-olds	0.73* (11/15)	—	—	—
Object-holding segment removed	3-year-olds	—	0.19** (3/16), Study 5	0** (0/15), Study 5	0* (0/15), Study 5
	5-year-olds	—	0.93** (14/15)	0.4 (6/15)	0.25* (4/16)

[a] If the number is smaller than .5 and has * or **, the object-same response was significantly above chance.

[b] The numerator represents the number of the action-same choosers; the denominator is the n in the condition.

*$p < .05$. **$p < .01$.

constructions to highlight the word's form class, Chinese 3- and 5-year-olds could not succeed. Three-year-olds still made AS choices marginally below chance (34.4%), $t(14) = -2.018$, $p = .063$, and 5-year-olds responded at chance (38.9%). Children's performance was no different from that in Study 4, in which the same monosyllabic verbs were presented in a single sentence frame. No age difference was found as well (see also Table 6 for the results of the individual analysis).

Results suggest that Chinese children as old as 5 years of age could not interpret a novel word as a verb and extend it to the same action. It appears that the lack of morphological distinction between nouns and verbs makes Chinese children map and extend novel verbs *less* rather than *more* readily. Furthermore, we suspected that the lack of morphological marking to distinguish the word's form class may lead Chinese children to rely relatively more heavily on *extralinguistic* cues.

One property of our stimuli may have given children a subtle cue that the object was the focus in the event. The standard video clips showed an actor holding an object for a moment (for about half a second) before starting the action. We shot the videos this way to make sure that children saw the object clearly. It should be stressed that the object was not unnaturally highlighted in the original stimuli, and it did not seem to affect Japanese- or English-speaking children. However, if Chinese children are particularly sensitive to extralinguistic cues because linguistic cues to form class are weaker than in Japanese or English, this first segment of the video might have lead Chinese children to think that the object was in a way "topicalized." We tested this possibility in Study 5.

Study 5: Can Chinese Preschoolers Map and Extend Novel Verbs When Actions Are Highlighted?

Would Chinese 3- and 5-year-olds map and extend novel verbs more readily when the action was highlighted by reducing the salience of the theme object? Even if we observe that the Chinese children perform better with the action-highlighted video, however, it may be because the action was made more prominent than the object. If true, children should select the action when the novel word is presented as a noun or a class-ambiguous word as well. To exclude this possibility, we also tested children in the noun and bare word conditions.

Method

Participants. Forty-six Chinese 3-year-olds ($M = 3.4$, range = 3.1–3.11) and forty-six 5-year-olds ($M = 5.4$, range = 4.9–5.11) were randomly assigned to the noun, bare word, or the verb with arguments condition. Children were normally developing, monolingual Mandarin-Chinese speakers living in Beijing. There were approximately equal numbers of boys and girls within each age and condition group, and none of them had been tested in the earlier studies.

Stimulus materials and procedure. The brief object-holding segment of the video was removed so that the object was already in motion at the start of the video. Other than this manipulation, the videos were exactly the same as those used in earlier studies. The procedure, instructions, and words used for the verb with arguments condition were identical to that in Study 4A. Thus, we presented monosyllabic nonsense words with three different aspect-marking auxiliaries in order to highlight that the word was a verb. The procedure, the instructions, and the novel words used in the noun and bare word conditions were identical to the corresponding conditions in Study 3.

Results and Discussion

Removing the object-holding segment brought about a drastic change in Chinese children's novel verb learning. In sharp contrast to the failure in the earlier studies, Chinese children's verb-learning performance was now equivalent to the level of performance by Japanese- or English-speaking children. Chinese 3-year-olds were now at the chance level, just like Japanese- and English-speaking 3-year-olds (40% AS responding). The 5-year-olds now made AS selections above chance level, just like their Japanese- and English-speaking age counterparts (88%), $t(15) = 9.934$, $p < .001$. On the individual analysis, 93% of the 5-year-olds were AS choosers, $p < .01$ (see Table 6).

Children's performance in the verb with arguments condition was compared with that in Study 4A, in which the same linguistic cues were provided but the initial object-holding section was present in the video clip. Results indicated that the children performed better with the videos in which the object-holding segment was removed than with the previous version of the video, $F(1, 57) = 11.91$, $p < .01$, and the magnitude of the gain by the video manipulation was larger for the 5-year-olds than for the 3-year-olds, as suggested by the significant Age × Video Type interaction, $F(1, 57) = 7.16$, $p < .01$.

Given that the manipulation of the videos made a dramatic difference in the performance in the verb with arguments condition, it is critical to see whether the children were influenced by this manipulation only in the verb learning context or whether their AS choices also increased in the noun and bare word conditions. In those conditions, AS choices should not predominate. Both 3- and 5-year-olds mapped the novel noun to the same-object video significantly above chance level, 67% and 73% OS responding, $t(15) = 2.44$, $p < .05$, and $t(14) = 4.58$, $p < .01$, respectively, and there was no difference across the two age groups, $F(1, 29) = .36$, $p > .1$. Thus, when children made AS selections in the verb with arguments condition, they did not do so solely because the action was made perceptually more salient than the object. However, when we compared the performance of the noun condition in this study to that in Experiment 3, there was a significant decrease in the proportion of OS responding, $F(1, 51) = 7.76$, $p < .01$, with no main effect or interaction involving age. The 3-year-olds in the bare word condition still mapped the class-ambiguous word to the object, making OS choices at test 72.2% of the time, $t(14) = 7.14$, $p < .01$, and the 5-year-olds did so 63.3% of the time, $t(14) = 1.26$, $p > .1$. The children's performance here was not different from their performance in the bare word condition with the original stimuli in Study 3, $F(1, 50) = 2.80$, $p = .1$.

When we compared the children in the bare word condition and those in the verb with arguments condition on the rate of AS responding, the children in the latter condition made AS choices much more frequently, $F(1, 57) = 56.10$, $p < .01$, but this was qualified by the Age × Condition interaction, $F(1, 57) = 20.08$, $p < .01$: The difference between the two conditions was highly significant in 5-year-olds, $F(1, 28) = 20.72$, $p < .01$, but not in 3-year-olds, $F(1, 29) = 2.34$, $p > .1$.

In summary, when the action was made salient by removing a short object-holding segment, Chinese 5-year-olds were able to map and extend novel verbs to the same action. However, this shift did not create an action-naming preference in the bare word case. Although this manipulation caused a significant decrease in the noun–object mapping, Chinese children still mapped novel nouns to the object rather than the action. These results suggest that Chinese children were affected by perceptual salience to some degree in novel verb learning as well as in noun learning but that the influence of this extralinguistic cue is much stronger for the former case.

The fact that the Chinese children performed on a par with their Japanese and English same-age peers in mapping and extending novel verbs allows us to rule out another artifact. Chinese children's prior performance in the earlier studies cannot be blamed on the fact that the task was unnatural or the instructions were inappropriate for Chinese children.

The removal of the brief object-holding segment from the original stimuli strongly helped Chinese 5-year-olds. Would this manipulation also help Japanese- and English-speaking 3-year-olds, who could not map and extend novel verbs? Study 6 used the edited videos with Japanese and English 3-year-olds, who were at chance level in Studies 1 and 2.

Study 6: Can Japanese- and English-Speaking 3-Year-Olds Map and Extend Verbs When the Actions Are Highlighted?

Method

Participants. Fourteen Japanese 3-year-olds ($M = 3.6$, range = 3.0–4.0) from monolingual families living in a suburban city in the Greater Tokyo area and 16 American 3-year-olds (M = 3.6, range = 3.0–3.11) from monolingual English-speaking families living in the Greater Philadelphia area participated. The demographic and ethnic backgrounds of the children were the same as those in Study 1 (Japan) and Study 2 (United States), although none participated in earlier studies. One Japanese child (side bias) and two American children (parental interference and failure to cooperate) were also tested but excluded from the final sample. The gender distribution was approximately equal within each language.

Stimuli and procedure. The edited stimuli from Study 5 were used. Because the Japanese children in Study 1 seemed to prefer the bare verb instruction over the verb with arguments condition, we used the same procedure, instructions, and novel words as those in the bare verb condition in Study 1. On the same logic, the American children were tested with the same procedure, instructions, and novel words as those in the verb with arguments condition in Study 2.

Results and Discussion

Japanese-speaking children. As shown in Table 3, the Japanese children performed at chance, making AS selections 52.4% of the time. Of the 14 children, 6 (39%) children were AS choosers, which was not different from chance by the binomial criterion, $p > .1$ (Table 4). The performance of these children was compared to that of the Japanese 3-year-olds in the bare verb condition in Study 1. No difference was found between the two groups.

English-speaking children. English-speaking 3-year-olds also responded at chance (53.1%). Of the 16 children, 6 children were AS choosers, $p > .1$. Children's performance was not different from that of the English-speaking 3-year-olds in the verb with arguments condition in Study 2 (42.3%).

Thus, unlike the case with Chinese 5-year-olds, the removal of the brief object-holding segment from the original stimuli did not improve Japanese or English 3-year-olds' performance in mapping and extending novel verbs. Of course, the results do *not* suggest that Japanese- or English-reared children do not use extralinguistic cues *in general*, nor that they are less sensitive to extralinguistic cues than Chinese children. With the linguistic cues available to them that signal verbhood, however, Japanese- and English-reared children did not need to notice whether a half second object-holding segment preceded the action event. We elaborate on this point in the General Discussion.

General Discussion

The present research examined whether children learn novel nouns more readily than novel verbs across different languages, and how a language's linguistic properties influence novel verb learning. This article empirically evaluated these questions using a novel word-learning paradigm across three very different languages (English, Japanese, and Chinese), the latter two of which have been purported to favor the learning of verbs in the literature (e.g., Choi & Gopnik, 1995; Ogura, 2001; Tardif, 1996, 2006).

Evidence for a Potential Universal Noun Advantage

The finding that 3-year-olds in all three languages succeeded in novel noun learning but not in novel verb learning supports the view that children universally map and extend novel nouns more readily than novel verbs (Gentner, 1982; Golinkoff et al., 1992, 1996; Imai, Haryu, Okada, Li, & Shigematus, 2006; Imai et al., 2005). In our task, it was not enough for children to form a mapping between an action and a verb. To succeed, children needed to *extend* the novel verb to another exemplar, a stringent test of verb learning. Thus, the results tell us that 3-year-old children are able to map and extend a novel noun to new instances after a single word-referent pairing (e.g., Golinkoff et al., 1992; Haryu & Imai, 2002; Imai & Haryu, 2001, 2004). However, they are not able to do so as readily for novel verbs (e.g., Golinkoff et al., 1996).

Additionally, the pattern of these results from Chinese children suggests that, given the choice, young children expect novel objects to be named over actions (Golinkoff et al., 1992; Markman, 1989). The bare word condition in Chinese offered an ideal case to examine the object-naming bias as the equivalent of a linguistic projective test: What meaning would children give to a novel word presented in the absence of syntactic support? Chinese children strongly favored an object-name interpretation for a novel word. Indeed, this result was all the more impressive given that Chinese children mapped the bare word to the object even when the object was deemphasized in the action event (Study 5), and given that Chinese adults preferred to map the bare word to the action rather than the object (Study 3A). Taken together, the fact that nouns get extended more readily than verbs and the fact that children seem to have a bias to label objects corroborate prior claims for a universal noun advantage (e.g., Bornstein & Cote, 2004; Gentner, 1982; Imai et al., 2005; Kersten & Smith, 2002).

Verb Learning: Differences Across Languages

Intriguing differences emerged across language groups in what cues children used. Although Japanese- and English-speaking 5-year-olds succeeded in the verb mapping and extension task, they did so under different linguistic conditions.

Verb learning in Japanese. Replicating the results of Imai et al. (2005), 5-year-old Japanese children reliably extended a novel verb to the same action when the verb was presented with the arguments dropped, although they did not succeed when the verb was presented with its arguments. For Japanese speakers, it is natural to drop verb arguments when they can be inferred from context. From the Japanese point of view, it was obvious that the subject was the actor and the theme object was the novel object in the bare verb condition. It is possible that Japanese children were distracted by hearing this unnecessary information. At the least, Japanese children did not appear to gain from information about argument structure in this particular context, in which the subject and the object could have been easily inferred from the video, in sharp contrast to English-speaking 5-year-olds.

Verb learning in English. English-speaking 5-year-olds successfully mapped and extended the verb to the same action only when it was accompanied by its arguments. Argument dropping does not occur nearly as often in English as in Japanese or in Chinese. By age 5, children seem to be sensitive to this fact. When the novel verb was presented in an unusual structural form, even though it was marked by the progressive /ing/, and even though the arguments of the verb could have been inferred from observation, children failed. This reliance on argument structure is consistent with research demonstrating that English-reared children exploit the argument structure surrounding verbs to infer verb meaning (Fisher, 1996; Fisher & Song, 2006; Hirsh-Pasek & Golinkoff, 1996; Naigles, 1990, 1996).

Verb learning in Chinese. There has been a common assumption in the literature that learning an argument-dropping language confers an advantage on verb learning (Choi & Gopnik, 1995; Gentner, 1982; Tardif, 1996). Indeed, previous research suggests that Chinese children have more verbs in their early lexicons than do English children, for example (Tardif, 2006).

From this point of view, Chinese and Japanese children might have been expected to perform better than English-speaking children. Furthermore, Chinese children might have exceeded Japanese children because of the morphological simplicity of Chinese verbs (Tardif, 1996). In contrast, Chinese children did not perform better than Japanese- or English-speaking children. In fact, Chinese 3- and 5-year-olds both bluntly failed to map a novel verb to a novel action. Chinese children were extremely sensitive to subtle contextual cues when learning novel verbs for action events. Unless the object was deemphasized, Chinese 5-year-olds were not able to map a novel verb to the action.

One might conclude from these data that the Chinese children relied solely on perceptual cues to assign a novel word to a form class. This, however, was not the case. Using the edited videos that downplayed the salience of the object, Chinese children in Study 5 were still able to map novel nouns onto objects. Crucially, this outcome demonstrates that children are using *multiple cues* to determine word reference. At present, we can only speculate about why Chinese children demonstrated such a strong object-naming bias and why they were so sensitive to contextual cues in our task. Indeed, the linguistic cues that Chinese-speaking adults readily detect were overridden in the face of a subtle extralinguistic cue that is a half second of an object-holding scene before the start of the action. The key linguistic property that sets Chinese apart from Japanese and English is the lack of morphological distinction between nouns and verbs. Thus, unlike the case with Japanese or English, Chinese speakers cannot determine the grammatical form class of a word by inflectional morphological markings. Of course, linguistic cues such as aspect-marking auxiliaries that accompany verbs (e.g., "zai" for progressive, "le" for perfective) are salient cues to indicate verbhood for adults but they are not inherent to the word nor are they obligatory. The present research indicated that even the presence of "zai" right before the verb may not be a strong enough cue for verbhood for young children when the following word is *novel*, even though they understand the "zai + verb" construction for familiar verbs. It is possible that other auxiliaries such as "le" signal verbhood more strongly than "zai" and might lead children to identify a novel word as a verb sooner. However, this is an empirical issue that requires future investigation.

In any event, in Chinese, even though word order provides a cue for determining the form class of each word in the sentence, it is only probabilistic: Although the basic word order is subject-verb-object, there are other word orders (object-subject-verb, subject-object-verb, and verb-object-subject) that also appear in spoken language (Li et al., 1993). Thus, to identify a word's grammatical class, Chinese speakers have to coordinate semantic and grammatical cues, both at the global (such as argument structure) and at the local (such as auxiliary verb) levels, in "a complex system of mutual constraints" (Li et al., 1993, p. 193).

Chinese also has a large number of words that can be used as both nouns and verbs. In addition, as mentioned earlier, in Chinese, the notion of "word" is much less transparent than in English and in Japanese (cf. Chao, 1968). For Chinese children, then, more sophisticated linguistic knowledge may be required to identify a verb and its argument structure than for Japanese- and English-speaking children. Perhaps because linguistic cues are less easily accessible in Chinese, extralinguistic cues might be weighted more heavily in early verb learning.

Universally Shared Cognitive Factors and Language-Specific Linguistic Factors

The crosslinguistic results suggest that universally shared cognitive factors and language-specific linguistic factors both matter for early word learning. Universal biases, such as the bias to label objects, appear to characterize early word learning given the striking similarity in the developmental pattern across the three languages. However, language-specific factors come into play as well, influencing the ease with which children can map novel verbs to novel actions and factors they rely on in this process.

The pattern of crosslinguistic data here fits nicely with the emergentist coalition model (ECM) proposed by Hollich, Hirsh-Pasek, and Golinkoff (2000; Maguire, Hirsh-Pasek, & Golinkoff, 2006). The ECM makes several predictions, each of which was confirmed by the crosslinguistic data presented here. First, given the perceptual prominence of stable objects relative to relational concepts, children should universally assume that words label objects before they label actions (verbs) or properties (adjectives; Gentner, 1982; Waxman & Booth, 2003). This pattern bears out for a number of languages (e.g., Bornstein & Cote, 2004), such as English (this research; Waxman & Markow, 1995), Spanish (Waxman, Senghas, & Benveniste, 1997), and Japanese and Chinese (this research). The fact that Chinese-speaking children assumed that a class-ambiguous word was an object name rather than an action name provides additional evidence for this prediction.

Second, because children move from a reliance on perceptual to social to linguistic cues for learning words, children's movement away from the object bias to label actions will be a function of the properties of the input language. Languages that contain only subtle linguistic cues to form class should require more linguistic knowledge. In their place, children should evince continued reliance on social cues. Because Chinese children appear to have fewer clear linguistic cues to verbhood than Japanese- or English-reared children, the ECM would predict that they rely on linguistic cues later than children learning these other languages.

All in all, our results suggest that word-to-world mapping seems to be influenced by linguistic characteristics of the input language. Further, they call for a reconsideration of the linguistic factors that have been assumed to contribute to verb meaning acquisition. Morphological simplicity, for example, may not offer the toehold into verb learning once posited. Here, however, we underscore the need for future research with more fine-grained analysis of linguistic properties that might affect verb learning. For example, even though Japanese and Chinese are both argument-dropping languages, the situations and frequencies in which argument dropping occurs may differ across the two languages, in addition to differences in the morphological marking associated with verbhood: Chinese speakers may drop arguments only when the referents have been explicitly established by prior discourse, whereas arguments may be dropped in Japanese whenever the speaker thinks that the hearer can recover the arguments from context. If this is the case, it is important to examine how these potential difference influences verb learning in Japanese and Chinese children. Future research should also examine how semantic factors such as imagability and concreteness of meanings influences verb learning (Ma, Golinkoff, Hirsh-Pasek, Macronough, & Tardiff, 2008).

Discrepancy Between Our Results and Results From Production Data and Checklists: The Demanding Nature of an Extension Task

Why did children as old as 3 fail in our task even though they comprehend and produce many verbs (e.g., Choi & Bowerman, 1991; Choi & Gopnik, 1995; Fenson et al., 1994; Golinkoff et al., 2001; Ma, Golinkoff, Song, & Hirsh-Pasek, 2008; Tardif, 1996)? Importantly, we are *not* claiming that verb learning does not occur until children are 5 years old. Early appearance of verbs in children's lexicon is universally evident. Yet, the appearance of verbs in the lexicon does not mean that children appreciate the full meanings of the verbs they use. Verb meanings develop progressively, and linguistic experience is necessary to finally obtain fully adult-like representations of verb meanings (Gallivan, 1988; Imai et al., 2005, 2006; Theakston et al., 2002).

Children's early verb meanings seem to be narrowly contoured. Children first seem to associate a novel causal action to a novel label at around 18 months of age (Casasola & Cohen, 2000). However, it is highly unlikely that 18-month-olds would be willing to extend a newly learned verb to the same action when one or both of the two objects (either the agent or the patient) are changed. Maguire et al. (2002) reported that English-speaking 18-month-olds were not willing to extend a newly learned verb to the same intransitive action when the actor was changed even after they had been trained on the verb–action association with four different agents. In fact, Maguire, Hirsh-Pasek, Golinkoff, and Brandone (2008) report that children at 30–38 months do best on extending a novel verb after viewing a single agent; they cannot extend a new verb after seeing multiple agents. The bulk of the evidence suggests that 3-year-olds are reluctant to extend a novel verb to new exemplars (e.g., Behrend, 1995; Forbes & Farrar, 1995; Imai et al., 2005; Kersten & Smith, 2002).

In summary, although children may be able to associate a novel verb with a novel action as early as 18 months, their representation of verb meanings is incomplete, as revealed by their failure to extend a newly introduced verb to a situation to which it is obviously extendable for adults. Extension is an excellent test of whether a child has truly learned a verb meaning because it requires flexibility and generative use. Thus, although 2-year-olds may say a verb, it may not be until much later that they come to possess adult-like representations of the verb (see also Imai et al., 2006). Our task required verb mapping and extension. Thus, given the task demands, children should reveal proficiency later than on checklists or from production data.

Why Is Verb Learning So Slow?

This article leaves us with an unresolved question: Why are children not willing to extend verbs, and why is verb learning so slow? We alluded to perceptual and conceptual factors: Objects are easier than actions to parse from the environment (Gentner, 1982; Gentner & Boroditsky, 2001; Maguire et al., 2006). Furthermore, although objects are temporally stable, actions are ephemeral. Object boundaries are easy to find. In contrast, the beginning and the ending of an action is not clearly marked, and the action itself often undergoes change as it proceeds. Our prior work (Imai et al., 2005; Study 4) indicates that children encode and remember objects more easily than actions. On the first day, children watched each of the six original videos for 30 s in the absence of verb labels. The next day, when children were asked to say which videos they had seen before, either the object or the action was changed. Even without language, Japanese 3-year-olds experienced more difficulty encoding and remembering novel actions than encoding and remembering novel objects. The fact that actions are less easily encoded than concrete physical entities (Gentner, 1982; Golinkoff et al., 1996; Maguire et al., 2006) may well contribute to the slowness of verb learning.

Yet, difficulty in encoding actions per se cannot be the sole reason for the slowness of full mastery of verb meanings. To extend a newly learned verb, children must learn the *semantic criteria* for generalization. An action can be done with many different objects. For example, we can throw a ball, a Frisbee, a stone, a disk, or almost anything we can lift up with our hands. Suppose that a child hears the verb "throw" while watching her father throw a ball at one time and sees her father lightly *toss* a ball at another time. At still a different time, the child sees a boy throw a Frisbee to a dog. The second scene is very similar to the first because the objects—the father and a ball—are the same. To fully learn the verb "throw," she needs to not only extend the word to the similar scene but also to an event in which the actors and objects are different. In other words, to extend a verb to other instances, children need to know that agents and objects can change as long as the action is the "same." Finding the invariants in the action across exemplars is discovering an action's "verbal essence" (Golinkoff et al., 2002). The mastery of this concept is a first step toward adult-like verb meanings and it is long in coming (Behrend, 1995; Forbes & Farrar, 1995).

We are not arguing, however, that 3-year-olds never generalize verbs when the patient object is changed. Perceptual salience as well as relative novelty between the object and the action in a given action event may well influence the ease with which children extract the invariants of the action from the whole scene. Given that there is no metric for measuring relative salience or novelty of objects and actions, we cannot conclude that our results are generalizable to any situation in which children must infer the meaning of a novel verb from observing an action event. In fact, in a different series of studies, we found that children's performance in verb generalization tasks improves when the changed patient object was perceptually similar to the original object or a familiar object (Haryu, Imai, & Okada, 2003; Imai, Haryu, & Okada, 2003; see also Kersten & Smith, 2002). Also, young children may extend a verb to a new exemplar (e.g., the same action with a new actor and/or new object), when they are trained with multiple different objects and/or with linguistic contrasts (cf. Bernal, Christophe, Waxman, Braun, & Lidz, 2005). However, the story is complicated. Children may need to have a firm representation of an action, possibly even in a specific context, before they can profit from seeing multiple, more varied exemplars (Maguire et al., 2008). At some point, however, multiple exposures must be essential for verb learning.

Coming to the insight that an action can be carried out with many different objects is only the first step for verb meaning generalization. In addition, word learners need to discover that different types of verbs employ different criteria for extension. Some verbs should be extended on the basis of sameness of manner, whereas others should be extended solely on the sameness of goals or results. Still others should be extended on the basis of the sameness of direction or trajectory of motion. Which semantic component is most likely to be packaged into verb meanings also differs across languages; therefore, children need to learn this distributional pattern (e.g., Choi & Bowerman, 1991; Talmy, 1985).

In the end, the question of whether noun learning is easier than verb learning may not prove to be the central question. The real issue is identifying the universal cognitive constraints on event interpretation and studying how these constraints and the semantic and syntactic structure of a lexical category in a particular language interact with the processes used to learn words. Given the structure of the noun (object names) lexicon, fast-mapping and extension with a single instance generally works fairly well, enabling children to build up a sizable vocabulary quickly. Given the complexity involved in the process of extracting the criteria for extension of a verb, it may be more adaptive to learn and extend verbs slowly, only after hearing the same verb used to describe similar actions with different objects across different situations.

Conclusions

This article has accumulated evidence informing three issues in language acquisition. First, it has complemented the evidence found with checklists and production data in showing that nouns are mapped and extended more readily (over verbs) across three languages with different linguistic properties. Second, because this is a crosslinguistic study employing the same paradigm, the data reveal that verb learning is a protracted process very much influenced the linguistic properties of the child's language. Finally, although perceptual, social, and linguistic cues are always available to verb learners, we have discovered that when a language offers minimal linguistic cues for identifying the form class membership of a novel word, children rely for a longer time on social cues.

References

Au, T. K.-F., Dapretto, M., & Song, Y. K. (1994). Input vs. constraints: Early word acquisition in Korean and English. *Journal of Memory and Language*, *33*, 567–582.

Behrend, D. A. (1995). Processes involved in the initial mapping of verb meanings. In M. Tomasello & W. E. Merriman (Eds.), *Beyond names for things: Young children's acquisition of verbs* (pp. 251–273). Hillsdale, NJ: Lawrence Erlbaum.

Bernal, S., Christophe, A., Waxman, S. R., Braun, I., & Lidz, J. (2005). *Syntactic and semantic cues to verb acquisition in 23 month-old infants: Evidence from French and English.* Poster presented at the biennial meeting of the Society of Research in Child Development, Atlanta, GA.

Bornstein, M., & Cote, L. R. (2004). Cross-linguistic analysis of vocabulary in young children: Spanish, Dutch, French, Hebrew, Italian, Korean, and American English. *Child Development*, *75*, 1115–1139.

Bowerman, M. (1980). The structure and origin of semantic categories in the language learning child. In M. L. Foster & S. H. Brandes (Eds.), *Symbol as a sense: New approaches to the analysis of meaning*. New York: Academic Press.

Brown, P., Gentner, D., & Braun, K. (2005, July). *Early acquisition of nouns and verbs in Tzeltal*. Paper presented at the 10th International Conference for the Study of Child Language, Berlin, Germany.

Casasola, M., & Cohen, L. B. (2000). Infants' association of linguistic labels with causal actions. *Developmental Psychology*, *36*, 155–168.

Chao, Y. R. (1968). *A grammar of spoken Chinese*. Berkeley, CA: University of California Press.

Childers, J., & Tomasello, M. (2002). Two-year-olds learn novel nouns, verbs, and conventional actions from massed or distributed exposures. *Developmental Psychology*, *37*, 698–705.

Choi, S. (2000). Caregiver input in English and Korean: Use of nouns and verbs in book-reading and toy-play contexts. *Journal of Child Language*, *27*, 69–96.

Choi, S., & Bowerman, M. (1991). Learning to express motion events in English and Korean: The influence of language-specific lexicalization patterns. *Cognition*, *42*, 83–121.

Choi, S., & Gopnik, A. (1995). Early acquisition of verbs in Korean: A cross-linguistic study. *Journal of Child Language*, *22*, 497–529.

Dromi, E. (1987). *Early lexical development*. Cambridge, UK: Cambridge University Press.

Erbaugh, M. (1992). The acquisition of Mandarin. In D. S. Slobin (Ed.), *The cross-linguistic study of language acquisition* (Vol. 3, pp. 373–456). Hillsdale, NJ: Lawrence Erlbaum.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D., & Pethik, S. J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, *59* (Serial Number 242).

Fisher, C. (1996). Structural limits on verb meaning: The role of analogy in children's interpretations of sentences. *Cognitive Psychology*, *31*, 41–81.

Fisher, C., & Song, H. (2006). Who's the subject? Sentence structure and verb meaning. In K. Hirsh-Pasek & R. M. Golinkoff (Eds.), *Action meets word: How children learn verbs* (pp. 392–425). New York: Oxford University Press.

Forbes, J., & Farrar, M. (1995). Learning to represent word meaning: What initial training events reveal about children's developing action verb concepts. *Cognitive Development, 10*, 1–20.

Forbes, J., & Poulin-Dubois, D. (1997). Representational change in young children's understanding of familiar verb meaning. *Journal of Child Language, 24*, 389–406.

Gallivan, J. (1988). Motion verb acquisition: Development of definitions. *Perceptual & Motor Skills, 66*, 979–986.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. A. Kuczaj (Ed.), *Language development: Vol. 2. Language, thought, and culture* (pp. 301–334). Hillsdale, NJ: Erlbaum.

Gentner, D., & Boroditsky, L. (2001). On individuation, relativity, and early word learning. In M. Bowerman & S. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 257–283). Cambridge, UK: Cambridge University Press.

Gentner, D., & Boroditsky, L. (2005, April). *Noun and verbs in early Navajo acquisition.* Paper presented at the biennial meeting of the Society for Research in Child Development, Atlanta, GA.

Gleitman, L. (1990). Structural sources of verb meaning. *Language Acquisition, 1*, 3–55.

Golinkoff, R., Chung, H., Hirsh-Pasek, K., Liu, J., Bertenthal, B., & Brand, R. (2002). Young children can extend motion verbs to point-light displays. *Developmental Psychology, 38*, 604–615.

Golinkoff, R., Hirsh-Pasek, K., Bailey, L. M., & Wenger, N. R. (1992). Young children and adults use lexical principles to learn new nouns. *Child Development, 28*, 99–108.

Golinkoff, R., Hirsh-Pasek, K., & Schweisguth, M. (2001). A reappraisal of young children's knowledge of grammatical morphemes. In J. Weissenborn & B. Höhle (Eds.), *Approaches to bootstrapping: Phonological, lexical, syntactic and neurophysiological aspects of early language acquisition.* Amsterdam: John Benjamins.

Golinkoff, R., Jacquet, R. C., Hirsh-Pasek, K., & Nandakumar, R. (1996). Lexical principles may underlie the learning of verbs. *Child Development, 67*, 3101–3119.

Haryu, E., & Imai, M. (2002). Reorganizing the lexicon by learning a new word: Japanese children's inference of the meaning of a new word for a familiar artifact. *Child Development, 73*, 1378–1391.

Haryu, E., Imai, M., & Okada, H. (2003, April). *The role of object in young children's understanding of action verb meaning.* Paper presented at the biennial meeting of the Society for Research in Child Development, Tampa, FL.

Hirsh-Pasek, K., & Golinkoff, R. (1996). *The origins of grammar.* Cambridge, MA: MIT Press.

Hollich, G., Hirsh-Pasek, K., & Golinkoff, R. (2000). Breaking the language barrier: An emergentist coalition model for the origins of word learning. *Monographs for the Society for Research in Child Development, 65*(3, Serial No. 262).

Huttenlocher, J., Smiley, P., & Charney, R. (1983). Emergence of action categories in the child: Evidence from verb meanings. *Psychological Review, 90*, 72–93.

Imai, M., & Gentner, D. (1997). A cross-linguistic study on constraints on early word meaning: Linguistic influence vs. universal ontology. *Cognition, 62*, 169–200.

Imai, M., Gentner, D., & Uchida, N. (1994). Children's theories of word meanings: The role of shape similarity in early acquisition. *Cognitive Development, 9*, 45–75.

Imai, M., & Haryu, E. (2001). Learning proper nouns and common nouns without clues from syntax. *Child Development, 72*, 787–802.

Imai, M., & Haryu, E. (2004). The nature of word learning biases and their roles for lexical development: From a cross-linguistic perspective. In D. G. Hall & S. R. Waxman (Eds.), *Weaving a lexicon* (pp. 411–444). Cambridge, MA: MIT Press.

Imai, M., Haryu, E., & Okada, H. (2003, November). *The role of argument structure and object familiarity in Japanese children's verb learning.* Paper presented at the 28th Annual Boston Conference on Language Development, Boston University, Boston.

Imai, M., Haryu, E., & Okada, H. (2005). Mapping novel nouns and verbs onto dynamic action events: Are verb meanings easier to learn than noun meanings for Japanese children? *Child Development, 76*, 340–355.

Imai, M., Haryu, E., Okada, H., Li, L., & Shigematus (2006). Revisiting the noun-verb debate: A cross-linguistic comparison of novel noun and verb learning in English-, Japanese- and Chinese-speaking children. In K. Hirsh-Pasek & R. Golinkoff (Eds.), *Action meets word: How children learn verbs* (pp. 450–476). New York: Oxford University Press.

Imai, M., & Mazuka, R. (2003). Re-evaluation of linguistic relativity: Language-specific categories and the role of universal ontological knowledge in the construal of individuation. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the issues of language and thought* (pp. 430–464). Cambridge, MA: MIT Press.

Imai, M., & Mazuka, R. (2007). Revisiting language universals and linguistic relativity: Language-relative construal of individuation constrained by universal ontology. *Cognitive Science, 31*, 385–413.

Kersten, A. W., & Smith, L. (2002). Attention to novel objects during verb learning. *Child Development, 73*, 93–109.

Kim, M., McGregor, K. K., & Thompson, C. K. (2000). Early lexical development in English- and Korean-Speaking children: Language-general and language-specific patterns. *Journal of Child Language, 27*, 225–254.

Landau, B., Smith, L., & Jones, S. (1988). The importance of shape in early lexical learning. *Cognitive Development, 3*, 299–321.

Lee, J., & Naigles, L. (2008). *Mandarin learners use syntactic bootstrapping in verb acquisition. Cognition, 106*, 1028–1037.

Li, P., Bates, E., & MacWinney, B. (1993). Processing a language without inflections: A reaction time study of sentence interpretation in Chinese. *Journal of Memory and Language, 32*, 169–192.

Li, P., Jin, Z., & Tan, L. H. (2004). Neural representations of nouns and verbs in Chinese: An fMRI study. *Neuroimage, 21*, 1533–1541.

Ma, W., Golinkoff, R. M., Hirsh-Pasek, K., Macronough, C., & Tardif, T. (2008). *Imageability predicts age of acquisition of verbs in Chinese children*. Manuscript submitted for publication.

Ma, W., Golinkoff, R. M., Song, L., & Hirsh-Pasek, K. (2008) *Chinese children's comprehension of extensions of familiar verbs*. Manuscript submitted for publication.

Maguire, M., Hennon, E., Hirsh-Pasek, K., Golinkoff, R., Slutzky, C., & Sootsman, J. (2002). Mapping words to actions and events: How do 18-month-olds learn a verb? In B. Skarabela, S. Fish, & A. H. Do (Eds.), *Proceedings of the 27th Annual Boston University Conference on Language* (pp. 371–382). Somerville, MA: Cascadilla Press.

Maguire, M., Hirsh-Pasek, K. Golinkoff, R. (2006). A unified theory of word learning: Putting verb acquisition in context. In K. Hirsh-Pasek & R. M. Golinkoff (Eds.), *Action meets word: How children learn verbs* (pp. 364–391). New York: Oxford University Press.

Maguire, M., Hirsh-Pasek, K., Golinkoff, R. M., & Brandone, A. (2008). Focusing on the relation: Fewer exemplars facilitate children's initial verb learning and extension. *Developmental Science*, *11*, 628–634.

Markman, E. M. (1989). *Categorization in children: Problems of induction*. Cambridge, MA: MIT Press.

Naigles, L. (1990). Children use syntax to learn verb meanings. *Journal of Child Language*, *17*, 357–374.

Naigles, L. (1996). The use of multiple frames in verb learning via syntactic bootstrapping. *Cognition*, *58*, 221–251.

Ogura, T. (2001). *Meishi yuui, doushi yuui ni oyobosu hahaoya no gengo nyuryoku no kentou* [Influence of maternal input in the distribution of nouns and verbs in early vocabulary] Ministry of Education and Science grant report. Kobe University, Japan.

Soja, N., Carey, S., & Spelke, E. (1991). Ontological categories guide young children's inductions of word meanings: Object terms and substance terms. *Cognition*, *38*, 179–211.

Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), *Language typology and syntactic description* (Vol. 3, pp. 57–149). New York: Cambridge University Press.

Tardif, T. (1996). Nouns are not always learned before verbs: Evidence from Mandarin speakers' early vocabulary. *Developmental Psychology*, *32*, 492–504.

Tardif, T. (2006). But are they really verbs? Chinese words for action. In K. Hirsh-Pasek & R. M. Golinkoff (Eds.), *Action meets word: How children learn verbs* (pp. 477–498). New York: Oxford University Press.

Tardif, T., Gelman, S., & Xu, F. (1999). Putting the "noun bias" in context: A comparison of English and Mandarin. *Child Development*, *70*, 620–635.

Theakston, A. L., Lieven, E. V. M., Pine, J. M., & Rowland, C. F. (2002). Going, going, gone: The acquisition of the verb 'go.' *Journal of Child Language*, *29*, 783–811.

Waxman, S. R., & Booth, A. (2003). The origins and evolution of links between word learning and conceptual organization: New evidence from 11-month-olds. *Developmental Science*, *6*, 128–135.

Waxman, S. R., & Markow, D. B. (1995). Words as invitations to form categories: Evidence from 12 to 13-month-old infants. *Cognitive Psychology*, *29*, 257–302.

Waxman, S. R., Senghas, A., & Benveniste, S. (1997). A cross-linguistic examination of the noun-category bias: Its existence and specificity in French- and Spanish-speaking preschool-aged children. *Cognitive Psychology*, *32*, 183–218.

Yamashita, Y. (1997). The effect of saliency in input on the acquisition of nouns and verbs: Innate constraints or language-specific input? *Naruto Eigo Kyoiku* (Naruto College of Education Research Papers), *11*, 177–188.

Appendix

Lists of Nonsense Words Used for Experiments 1–7

	1	2	3	4	5	6
Japanese	neke(-tteiru)[a]	yachi(-tteiru)	chimo(-tteiru)	nuhe(-tteiru)	ruchi(-tteiru)	heku(-tteiru)
English	dax(ing)[b]	jav(ing)	blick(ing)	twill(ing)	glorp(ing)	modi(ing)
Chinese bisyllabic	dang1[c]pa4	wan4cu4	fa2ku1	suo3han4	mou2da2	wa1mo4
Chinese monosyllabic	jing2	gui2	jiong4	kuo2	lia1	pan3

[a] The morpheme in the parenthesis is the one marking the progressive aspect when the word was presented as a verb.
[b] -ing was added when the word was presented as a verb.
[c] The numbers indicate the tone corresponding to the morpheme.

Child Development, July/August 2008, Volume 79, Number 4, Pages 1001–1015

Does Chronic Classroom Peer Rejection Predict the Development of Children's Classroom Participation During the Grade School Years?

Gary W. Ladd, Sarah L. Herald-Brown, and Mark Reiser
Arizona State University

A sample of 398 children was followed up from ages 5 to 12 to investigate the relation between peer group rejection and classroom participation. The participation trajectories of individuals and groups of children who were rejected for differing periods of time were examined both during and after rejection using piecewise growth curve analyses. The results showed that whereas during periods of rejection, children exhibited negative or negligible growth in participation, when nonrejected, they manifested positive growth. These findings corroborated the hypothesis that (a) peer rejection creates constraints that inhibit children's classroom participation and (b) the cessation of rejection enables children to become more active and cooperative participants in classroom activities.

Support for the premise that classroom peer relations shape children's adjustment to school has grown. Nearly 20 years ago, Parker and Asher (1987) analyzed extant evidence and concluded that low peer acceptance predicted school maladjustment, particularly during adolescence. Subsequent evidence has corroborated this conclusion and shown that poor peer relations antecede school adjustment problems at earlier stages of children's school careers (Ladd, 2005; MacDougall, Hymel, Vaillancourt, & Mercer, 2001; Perry & Weinstein, 1998).

To elucidate this linkage, investigators have examined multiple features of children's classroom peer relations (e.g., friendship, friendship quality, peer group acceptance/rejection, and peer victimization; Ladd, 2003, 2005; MacDougall et al., 2001; Parker, Rubin, Price, & DeRosier, 1995) as predictors of their school adjustment. Results indicate that classroom peer rejection—which has typically been defined as how consensually disliked (relative to how consensually liked) a child is by members of his or her peer group (see Bukowski & Hoza, 1989)—is one of the strongest predictors of children's school adjustment. For example, Vandell and Hembree (1994) found that even after controlling for friendships and family characteristics, children who were rejected by peers were prone to develop scholastic difficulties. After adjusting for the predictive contributions of friendship and peer victimization, Ladd, Kochenderfer, and Coleman (1997) found that peer group acceptance/rejection was the best relational predictor of changes in kindergartners' classroom involvement and academic progress. Furthermore, Ladd and colleagues found that classroom peer rejection predicted some indicators of school maladjustment not only as children negotiated the challenges of kindergarten (e.g., school avoidance, negative school attitudes, and poor achievement; Buhs & Ladd, 2001; Ladd, 1990; Ladd, Birch, & Buhs, 1999) but also as they progressed through the early elementary school years (Buhs, Ladd, & Herald, 2006; Ladd & Burgess, 2001).

These longitudinal findings illustrate that even during the early grade school years, classroom peer rejection antecedes multiple forms of school adjustment, including children's school attitudes, school engagement, and scholastic achievement (Ladd, 2003). Although all these linkages warrant further investigation, the predictive significance of peer rejection for children's classroom participation may be among the most important to explicate. As Finn (1993) has argued, it is possible to conceptualize classroom engagement in behavioral terms and index facets of this construct by measuring the extent and quality of children's participation in classroom activities. It has

---

This investigation was conducted as part of the Pathways Project, a larger longitudinal investigation of children's social/psychological/scholastic adjustment in school contexts that is supported by the National Institutes of Health (1 RO1MH-49223, 2-RO1MH-49223, and R01HD-045906 to Gary W. Ladd). Special appreciation is expressed to all the children and parents who made this study possible and to members of the Pathways Project for assistance with data collection. Portions of this study were reported in a master's thesis that was completed by Sarah L. Herald-Brown at Arizona State University, and the authors gratefully acknowledge committee members Becky Kochenderfer-Ladd and William Fabricius for their cogent contributions.

Correspondence concerning this article should be addressed to Gary W. Ladd or Sarah L. Herald-Brown, P.O. Box 853701, School of Social and Family Dynamics, Arizona State University, Tempe, AZ 85287-3701. Electronic mail may be sent to gary.ladd@asu.edu or to sarah.herald@asu.edu.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7904-0013

been hypothesized that children must cultivate two forms of classroom participation—termed cooperative and independent participation—to maximize learning and achievement (Finn, 1989, 1993; Fredricks, Blumenfeld, & Paris, 2004; Ladd, Buhs, & Seid, 2000; Wentzel, 1991). Whereas independent participation denotes children's propensity to take initiative (i.e., independently seek out and pursue classroom activities), cooperative participation refers to children's adoption of the "student role" and compliance with classroom social rules and expectations (i.e., following directions, working on assigned tasks, adhering to school and classroom rules, etc.; Buhs & Ladd, 2001; Wentzel, 1991). Evidence assembled thus far corroborates the premise that both forms of participation forecast many aspects of children's school adjustment, including their scholastic progress (see Buhs & Ladd, 2001; Buhs et al., 2006; Finn, 1989; Ladd et al., 1999).

Accordingly, the purpose of this investigation was to further elucidate the relation between classroom peer rejection and the development of independent and cooperative classroom participation during the grade school years (kindergarten to Grade 6). The framework that guided this investigation was based on theory and research on the role of rejection-related processes in the development of children's school adjustment (Buhs & Ladd, 2001; Buhs et al., 2006; Coie, 1990; Ladd, 2003). Key tenets within this perspective include the following: (a) children who become collectively disliked (i.e., rejected) by classmates are increasingly subjected to peer-imposed maltreatment, including interpersonal constraints (e.g., ignoring, avoidance, and exclusion) and abuse (i.e., ridicule, harassment, and physical harm), and (b) exposure to, and especially, *sustained* experience with these rejection processes not only restricts children's opportunities to participate in classroom activities but also discourages productive forms of classroom engagement (e.g., reduces children's motivation to seek out learning activities, interact cooperatively with classmates).

The first of these two propositions has been supported by evidence indicating that peers tend to ignore, avoid, exclude, abuse, or otherwise limit the quality and quantity of rejected children's participation within classroom and playground activities, including those designed to promote learning and scholastic competence (Asher, Rose, & Gabriel, 2001; Buhs & Ladd, 2001; Ladd, Price, & Hart, 1990). The second proposition is the more innovative and least researched of the two and is based on the assumption that exposure to rejection processes expedites maladaptation, and the longer it is sustained, the greater the toll it takes on children's adjustment (Ladd, 2003). Although not as extensive, findings pertinent to this tenet indicate that sustained rather than transient classroom peer rejection (i.e., chronic peer rejection) increases the likelihood that children will develop adjustment problems (DeRosier, Kupersmidt, & Patterson, 1994; Ladd & Burgess, 2001; Ladd & Troop-Gordon, 2003). For example, DeRosier et al. (1994) found that children who experienced longer periods of rejection exhibited more serious psychological difficulties. Burks, Dodge, and Price (1995) also found that longer periods of rejection predicted increased risk for internalizing problems, but their results were limited to boys. Similarly, Ladd and Troop-Gordon (2003) established that chronic exposure to adverse peer relationships was a more powerful risk factor than strains present in children's contemporary peer relationships.

These findings are important because they link the duration of peer rejection with the severity of children's adjustment problems. At present, however, evidence is limited in that it has more to say about how chronic rejection is related to children's psychological adjustment than about its association with classroom participation during the grade school years. Furthermore, insight into the latter linkage has been occluded by design limitations, such as the propensity to study this association concurrently or across a limited number of school years. Thus, what has been learned thus far is that relatively brief periods of chronic rejection during grade school correlate positively with indicators of school maladjustment either concurrently (cross-sectional data) or at a single subsequent time point (follow-up longitudinal evidence). Because it has been rare for investigators to study the association between chronic peer rejection and classroom participation prospectively across substantial periods of schooling, little is known about how chronically rejected children *progress* toward maladaptive outcomes. Insight into this process could be gained by determining whether increments in the duration of chronic rejection predict more extended or accelerated patterns of classroom disengagement, and whether shorter versus longer periods of chronic rejection across the grade school years orient children toward differing participation trajectories.

Pursuit of these objectives also created the opportunity to explore new questions. A further implication of the premise that variability exists in the duration of rejection experiences is that children's exposure to this kind of relational adversity is not always continuous over the course of grade school. Recognizing that there also can be discontinuity in children's rejection experiences raises the question of what happens to children's classroom participation when their exposure to peer rejection ceases. To our knowledge, no one has investigated whether children's classroom participation trajectories remain stable or change after the offset of peer rejection and, in particular, whether shifts in participation trajectories are more probable or dramatic following brief versus sustained periods of peer rejection.

To address these objectives, we conducted a prospective, longitudinal investigation that spanned the elementary school years (kindergarten to Grade 6) and pursued two specific aims that were intended to elucidate the links between continuity/discontinuity in peer group rejection and children's classroom participation trajectories. Our first aim was based on the premise that exposure to peer group rejection inhibits the growth of classroom participation because rejection-related processes not only restrict children's opportunities to participate in classroom activities but also discourage productive forms of classroom engagement (e.g., reduces children's expression of initiative and their adoption of the student role in classroom activities). To be specific, we hypothesized that longer periods of chronic rejection inhibit children's classroom participation to a greater extent than shorter periods of chronic peer rejection.

The logic of this hypothesis is consonant with theories of psychological risk and stress (Dohrenwend & Dohrenwend, 1981; Johnson, 1988), in which it has been argued that the probability of maladjustment increases as a function of the length of time children are exposed to interpersonal stressors or risk factors. Furthermore, there is a modicum of evidence that implies that chronic rejection, more than brief or transient rejection, antecedes more prolonged or serious adjustment difficulties. DeRosier et al. (1994), for example, found that children who were rejected over a 2- to 3-year period, as compared to those rejected for 1 year or less, were at heightened risk for maladjustment. Evidence also links chronic rejection with *school adjustment* problems (Ladd & Burgess, 2001). Based on these findings, it was hypothesized that although children were experiencing chronic rejection, they would exhibit lower levels of classroom participation than nonrejected children, including significantly lower levels of independent and cooperative participation. Because the specific ways that chronic rejection might inhibit children's classroom participation have not been examined, we also explored whether longer periods of rejection were associated with (a) longer periods of stability or decline in participation trajectories, (b) more rapidly decelerating growth in participation trajectories, or (c) both of these possible temporal patterns.

As a first step toward addressing these hypotheses, we examined children's peer group acceptance/rejection trajectories across the grade school years and attempted to identify distinct groups whose patterns of movement in and out of rejection differed with respect to timing (i.e., earlier, later, or repeated onsets/offsets during grade school) and chronicity or duration (i.e., shorter vs. longer periods of rejection). Previous research on rejection chronicity (DeRosier et al., 1994; Ladd & Burgess, 2001; Ladd & Troop-Gordon, 2003) led us to expect that several types of rejection trajectories (i.e., rejection subgroups) might be identified, including (a) children who were rejected during the early but not during the later years of grade school, (b) children who were rejected in later but not in early years of grade school, (c) children who cycled in and out of rejection multiple times across grade school, (d) children who were rejected throughout grade school (stably rejected), and (e) children who never were rejected throughout grade school.

As part of this same aim, we examined the question of whether changes in peer group rejection—specifically, evidence of (a) a transition away from acceptance and into rejection or (b) a transition away from rejection and into acceptance among classmates—were associated with changes in children's classroom participation trajectories. Consistent with our initial premises, it was expected that the participation trajectories of children moving into rejection would decline or worsen over time. For children moving out of rejection, it was anticipated that participation trajectories would either (a) endure or worsen over time (continuity hypothesis) or (b) improve over time (discontinuity hypothesis). The possibility of these two postrejection trajectories was considered because no one has previously investigated whether children's participation remains stable or changes following the offset of peer group rejection and because both trajectories appear credible in light of extant theory and evidence.

The continuity hypothesis was based on the proposition that peer rejection suppresses children's initiative and their inclination to adopt the student role, and once inured, these participation trajectories endure or worsen as children progress through school (Buhs & Ladd, 2001; Finn, 1989; Wentzel, 1991). In contrast, the discontinuity hypothesis was built on the tenet that chronic peer rejection's effects on classroom participation trajectories are limited largely to the period during which children are experiencing rejection or rejection-related processes. Thus, in contrast to the continuity model, it could be expected that established nonparticipation trajectories would dissipate or reverse themselves after peer rejection ceased. Support for this postulate includes evidence suggesting that the effects of adverse peer experiences are not always permanent (Kochenderfer & Ladd, 1996; Kochenderfer-Ladd & Wardrop, 2001), and the assertion that children are more resilient at recovering from peer adversity during childhood than during later periods of their development (Ladd, Buhs, & Troop, 2002).

To test hypotheses about continuity/discontinuity in children's participation trajectories, it was necessary to identify groups of children whose acceptance/rejection trajectories fit previously mentioned patterns (a) and/or (b)—that is, groups that transitioned into or out of chronic rejection at one or more points during the kindergarten to grade school years. Our goal for all such groups was to compare their participation growth rates, using piecewise growth curve analyses, both before and after they were rejected. Moreover, regardless of the timing of rejection's onset, it was hypothesized that the duration of rejection (rejection chronicity) would be inversely related to growth in participation—that is, the longer children remained rejected, the less growth they were expected to exhibit in classroom participation. Support for the continuity model would consist of findings indicating that peer rejection orients children toward an enduring, irreversible disengagement trajectory—one that is maintained or becomes further elevated even *after* children are no longer rejected by their classmates. Conversely, corroboration of the discontinuity model would include findings showing that children's participation trajectories improve after the offset of chronic rejection—that is, the data pattern would imply that the cessation of peer rejection deflects children from their preexisting disengagement trajectories.

To increase the probability that observed differences in participation trajectories were attributable to variations in chronic peer rejection, data were analyzed to examine and, if necessary, control for potential covariates that might be associated with disparities in classroom participation. Because variables such as aggressive and withdrawn behavior, ethnicity, and family socioeconomic status (SES) have been linked with children's peer rejection and school adjustment (Ladd et al., 1999), these factors were examined as covariates.

The second aim of this investigation was to determine whether variations in chronic peer rejection were differentially predictive of children's cooperative versus independent participation trajectories. Although extant theory and evidence imply that longer periods of peer rejection impede the development of classroom participation, there is no comparable conceptual or empirical basis for formulating predictions about chronic rejection's effects on specific forms of classroom participation. Therefore, investigation of this aim was guided by rationales extrapolated from other sources, including research on rejected children's peer interactions, and educational theory and practice.

It seems likely that chronic classroom rejection is a detriment to the development of children's independent participation. Researchers have established that peers respond more negatively to rejected than to nonrejected children's initiatives (Dodge, 1983; Ladd et al., 1990; Putallaz & Gottman, 1981), and contemporary educational practice creates ample opportunities for classmates to negate rejected children's initiatives. Peer-mediated learning activities (e.g., cooperative learning groups, peer collaboration, activity-based instruction) have become increasingly prevalent in American grade schools (Damon & Phelps, 1989; Furman & Gavin, 1989; Radziszewska & Rogoff, 1988; Slavin, 1983), and these methods encourage classmates to respond to each other's initiatives (e.g., suggestions, requests, plans, and other expressions of initiative). If rejected children are persistently entrenched in classroom environments where their initiatives are ignored, ridiculed, or otherwise demeaned by classmates, then they may learn to refrain from such behaviors as a means of protecting themselves from further abuse. Furthermore, longer periods of rejection might inhibit independent participation because it causes children to develop feelings and cognitions that are antithetical to self-assertion and initiative (e.g., low self-efficacy, feelings of incompetence, learned helplessness). Hence, it was expected that longer exposures to peer rejection would reduce or suppress the growth of children's independent participation.

It is also conceivable that chronic rejection restrains the development of cooperative classroom participation. It has been proposed that adverse classroom experiences not only discourage children from identifying or "bonding" with school but also serve as an impetus for certain forms of behavioral disengagement, such as detachment from and resistance to the classroom culture (Finn, 1993). Because chronic rejection represents an enduring form of classroom adversity (i.e., sustained exposure to a rejecting peer environment), it may undermine school identification and encourage children to detach from the student role and disregard or resist classroom rules and expectations. However, the effect of chronic rejection on cooperative participation may not be as strong as its impact on independent participation because the former aspect of engagement is also shaped by many other socializers, some of whose influences might be more powerful than classmate's rejecting behaviors.

For example, teachers and school personnel expend considerable effort promoting cooperative participation during the early grades because it is viewed as an essential prerequisite for learning and achievement (Blumenfeld, Hamilton, Bossert, Wessels, & Meece, 1983; Wentzel, 1991). Similarly, parents often socialize these behaviors from an early age as a means of preparing children for school. For these reasons, it was anticipated that chronic rejection would be more strongly associated with changes in children's independent participation compared to cooperative classroom participation.

Method

Participants

Participants for this study were 398 children (199 boys and 199 girls) who were recruited into a longitudinal investigation as they entered kindergarten ($M_{\text{age}} = 5.50$, $SD = 0.34$) and followed yearly until they completed sixth grade ($M_{\text{age}} = 11.7$, $SD = 0.57$). Consent was obtained from school districts before recruitment began, and written informed consent was obtained from parents prior to children's participation in the study. Of the families who were invited to join the investigation, 95% consented to their child's participation. European American children (77.5%), African American children (17.9%), Latino/Hispanic children (1.3%), and children from mixed or other ethnic/racial backgrounds (3.3%) were included in the sample. At the time of recruitment, children lived in a wide range of socioeconomic households: 36.8% were lower to middle income ($0 – $20,000), 30.6% were middle income ($21,000 – $40,000), and 32.6% were middle to high income (above $41,000).

Over the course of this investigation, participants became increasingly dispersed across classrooms and school districts. When children moved or changed schools, permission was sought from administrators, teachers, and parents to extend the project into their schools. Only those classmates who had written informed parental consent (and child assent) participated in the study. The number of teachers who participated in this study per assessment period ranged from 32 to 282, and the number of classmates who contributed data per assessment ranged from 964 to 3,030.

Measures

To the extent possible, data were gathered from multiple informants to reduce problems associated with shared method variance. Different informants provided data on children's peer rejection (i.e., classmates), classroom participation (i.e., teachers), behavior with peers (i.e., independent observers), and family SES and ethnicity (i.e., parents).

Peer rejection. A roster-and-rating sociometric technique was administered in classrooms to obtain classmate's reports of children's peer acceptance/rejection (Asher & Dodge, 1986; Asher, Singleton, Tinsley, & Hymel, 1979). Classmates (those with permission) rated each other on the following likability criterion: "How much do you like to play with this person in school?" For fourth-grade and older participants, this criterion was changed to " . . . hang out with . . .."

Prior to administering the scale, participants were taught to define the scale points as 3 = *I like to play with this person a lot*, 2 = *I like to play with this person sometimes*, and 1 = *I don't like to play with this person very much*. For children in kindergarten to Grade 2, a felt board containing classmate's pictures was used instead of a class roster, and respondents rated classmates during individual interviews by sorting their pictures into boxes marked (in text and graphics) with the same three scale points. Because the definitions of scale points were somewhat lengthy, these anchors were abbreviated on sort boxes and class rosters as "a lot," "sometimes," and "not much." The ratings children received from classmates were averaged and standardized within classrooms. It has been established that this measure yields reliable and valid data with younger and older children (Asher et al., 1979; Cillessen & Bukowski, 2000; Ladd, Herald, & Andrews, 2006).

Independent and cooperative classroom participation. Measures of these constructs were obtained during the spring of every grade by asking teachers to rate items from corresponding subscales of the Teacher Rating Scale of School Adjustment (Birch & Ladd, 1997; Ladd, Kochenderfer, & Coleman, 1996). These subscales yield reliable and valid scores with grade school children (Buhs et al., 2006; Ladd et al., 2000). Independent participation was measured with a four-item subscale (sample items: "works independently," "self-directed child"; αs ranged from .85 to .89), and cooperative participation was assessed with a seven-item subscale (sample items: "accepts responsibility for a given task," "follows teacher's directions"; αs ranged from .88 to .93). All items were rated on a 3-point scale (0 = *doesn't apply*, 1 = *applies sometimes*, 2 = *certainly applies*), and ratings were averaged within subscales to obtain a score for each child.

Aggressive and withdrawn behavior with peers. Trained observers visited classrooms several times per week during the first 10 weeks of kindergarten and used a combination point-time and scan sampling technique (Altman, 1974; Ladd & Profilet, 1996) to sample the percentage of time participants spent in aggressive and withdrawn behaviors. Children were observed during free-play periods; after moving within earshot of a target child, observers first focused on a participant (2 – 3 s) and then looked away for an equal interval (2 – 3 s) to code the child's behavior into 1 of the 17 mutually exclusive behavioral categories. These on- and off-look intervals were repeated until a total of 27 scans had been recorded. Observers then located the next target child (children were observed in a predetermined, random order) and repeated this sampling procedure and so forth, until a set of 27 scans had been collected on all participants within a classroom. Once this cycle was completed, another set of scans was undertaken until a total of 20 sets of scans (a total of 540 scans per child) were gathered. These sets of scans were spaced at nearly equal intervals across the 10-week observation period.

Four of the 17 observational categories were used in this study: aggression, arguing, onlooking, and unoccupied. Behaviors coded as aggression and arguing with peers (i.e., hostile acts and hostile talk, respectively) were averaged to form an aggression score, and behaviors coded as onlooking and unoccupied (i.e., watching others but not interacting, and alone and off-task, respectively) were averaged to create a withdrawn behavior score.

Prior to data collection, observers learned children's names and practiced with a judge until they achieved a minimum of 80% agreement on all behavioral codes. Observers were then assigned to one or more classrooms and conducted 17% – 20% of their observations on a scan-by-scan basis with a reliability judge. Kappas for the entire scheme, which were calculated between each observers and judges, exceeded .70, and percent agreement on the codes used in this study averaged 86.8% (range = 73% – 100%).

Family SES. This variable was assessed with Entwisle and Astone's (1994) socioeconomic index (SEI). Adult caregivers (living in the home) received a demographic questionnaire as children began kindergarten and were asked to list their primary occupations and also write a description of their job and the duties it entailed. Data were gathered by phone or in person for caregivers who could not provide a written response. Participant's responses were coded by trained undergraduates into distinct occupational status categories and converted to corresponding SEI scores. Twenty percent of these data were scored by independent coders, and Kappa between pairs of coders on categories (agreement vs. disagreement) met or exceeded 78%. The highest of the SEI scores obtained within the same family was utilized as the participant's SEI score. Scores ranged from 0 to 97.16, with a median of 49.14 (scores around 50 are assigned to occupations such as electronic sales personnel, administrative support positions).

Ethnicity/race. Participants' parents indicated their children's ethnicity/race on the same demographic questionnaire that was used to gather SEI data. The categories used for this purpose were (a) European American, (b) African American, (c) Latino/Hispanic, (d) Native American, (e) Asian or Pacific Islander, and (d) other (other or mixed ethnic/racial backgrounds).

Procedure

From kindergarten through Grade 6 (for 7 years), data from participants, teachers, and parents were gathered during the spring of each school year, and within each assessment occasion all measures were administered in counterbalanced order. Participants completed the sociometric measure within individual interviews during the kindergarten, first, and second grades, and within group settings (e.g., classrooms) thereafter. This alteration reflects standard practice in research on children's peer acceptance and rejection. Observers gathered data on children's behavior, and parents completed a demographic questionnaire during the fall of children's kindergarten year. Teachers completed the classroom participation measures during each year of the project. Participants received age appropriate thank you gifts (i.e., colorful pencils, folders, and gift certificates), and teachers and parents received a cash honorarium.

Results

Data Analytic Strategy and Results of Preliminary Analyses

Variable distributions, relations, and patterns of missing data were examined to determine whether these and other properties of the data set conformed to the assumptions underlying parametric statistics and data augmentation procedures. Because it was possible to satisfy these prerequisites, multiple imputation was used to replace missing values with estimated scores. Multicollinearity and construct independence were evaluated by examining relations among predictors or criteria. Confirmatory factor analyses (CFAs) were used to evaluate dimensionality and factorial invariance of the participation measures.

The hypothesized differences in participation trajectories for individuals and groups of chronically rejected and nonrejected children were examined using piecewise growth curve analyses.

Estimation of Missing Data via Multiple Imputation

Attrition was minor with over 95% of the sample participating from kindergarten to Grade 6. Constraints on data collection (e.g., absences, moves, dropouts) caused 3.1% of the longitudinal data points to be missing (range by variable = 0.00% – 14.1%).

Missing data were estimated via NORM (Schafer, 1999), a multiple imputation program in which missing multivariate data are simulated $m > 1$ times. Three imputations were generated because for data sets containing 10% or fewer missing data points, it has been established that $n = 3$ imputations yield sufficient estimation efficiency (i.e., 97%; Rubin, 1987; Schafer, 1999). Mplus statistical software was used to estimate each of the hypothesized models from the multiple imputed data sets and to combine parameter estimates and standard errors across imputed data sets (Muthén & Muthén, 1998 – 2004).

Psychometric Properties and Factorial Invariance of the Participation Criteria

The items within the independent and cooperative participation subscales yielded scores that had no serious floor or ceiling effects and thus were retained as indicators of their designated subscales. Separate CFAs were performed for the independent and cooperative participation criteria to examine the factor structure and invariance of each measure over time. Each CFA examined the fit of a tau-equivalent model that contained seven latent participation variables—one per grade level (grade was the grouping variable)—each of which was constructed from the same set of indicators (i.e., the items administered for the subscale at each time of assessment). This model fit the data adequately for independent participation (comparative fit index [CFI] = 0.97, root mean square error of approximation [RMSEA] = 0.05, standardized root mean square residual [SRMR] = 0.05) and for cooperative participation (CFI = 0.94, RMSEA = 0.05, SRMR = 0.05), supporting the assumption that each measure tapped a single factor that was invariant (i.e., equivalent lambdas and error variances) across the seven times of assessment. Scores for independent and cooperative participation were created by averaging the ratings obtained for each subscale's items at each time of measurement (see the Method section for subscale alphas).

Relations Among Predictors and Criteria

Predictors were correlated by time of measurement, as were criteria. Multicollinearity among predictors (i.e., sex, SEI, ethnicity, aggression, and withdrawal) was not evident because the coefficients obtained were low to moderate in magnitude (*r*s ranged from −.32 to .44). In contrast, the participation criteria correlated positively and substantially across kindergarten through Grade 6 (*r*s ranged from .58 to .73). Hence, two CFAs were conducted to determine whether the independent and cooperative participation measures were sufficiently distinct to be evaluated as separate criteria. The first CFA evaluated the fit of a model in which items from both the independent and the cooperative participation measures served as indicators of a single latent variable at each time of measurement. The second CFA examined a two-variable model in which items served as indicators of their respective participation constructs at each time of assessment. Results showed that the single-latent-variable model, $\chi^2(36) = 333.72$, $p < .001$; Akaike information criterion (AIC) = 6,415.19, Bayesian information criterion (BIC) = 6,533.79, fit the data less well than did the two-latent-variable model, $\chi^2(26) = 113.98$, $p < .001$; AIC = 5,634.05, BIC = 5,748.69, and an examination of these information criterion indicated that the latter model fit the data significantly better than did the former one (i.e., smaller AIC and BIC values indicate better fit). Because these findings implied that different information was contained in the two participation measures, both were utilized in subsequent analyses.

Identifying Groups of Children With Distinct Kindergarten to Grade 6 Peer Rejection Histories

A series of analytic steps were taken to identify and describe groups of children whose rejection trajectories (i.e., patterns of movement in and out of rejection) across the kindergarten to Grade 6 school years were distinct with respect to timing (i.e., earlier, later, or repeated onsets/offsets during grade school) and chronicity (i.e., shorter vs. longer periods of rejection). First, for each participant in the sample, we determined which grades (if any) during the 7-year interval from kindergarten to Grade 6 that she or he received an averaged peer acceptance/rejection rating (from classmates) that was indicative of peer group rejection. Within each grade level, participants were designated as "rejected" if their average likability rating from classmates was more than two thirds of 1 *SD* below the mean likability rating for their classroom (i.e., their standardized peer rating score was <-0.70 *SD*; see Measures section). Averaged peer ratings rather than negative peer nominations were used for this purpose because low scores on this criterion provide a comparable marker of peer group rejection (Asher & Dodge, 1986) and because ratings possess superior psychometric properties (e.g., greater variability, normally distributed rather than skewed scores) and, thus, provided a more reliable index and consistent criterion for distinguishing rejected from nonrejected children across multiple occasions and classrooms. Although investigators sometimes use cutoffs of -1 *SD* to define rejection, reliance on less conservative criteria (e.g., $-.50$) is not uncommon. Because the former criterion identifies fairly extreme cases and would not identify children whose scores varied slightly from year to year, a moderately conservative criterion of $-.70$ was utilized (cf. Ladd & Troop-Gordon, 2003).

Second, we examined the number and sequence of grade levels across the kindergarten to Grade 6 school years that each participant met the criterion for peer group rejection, and on the basis of this information, each child in the sample was assigned to one of the following six mutually exclusive groups: (a) early chronic rejected (ECR) group (i.e., children rejected in three or four grades during the kindergarten to Grade 3 school years, but no more than one grade thereafter; $n = 45$); (b) later chronic rejected (LCR) group (i.e., children rejected in two or three grades during the Grades 3–6 school years, but no more than one grade prior to this; $n = 33$); (c) 1-year rejected (1-YR) group (i.e., children rejected in only one grade during the kindergarten to Grade 6 school years; $n = 56$); (d) sporadically rejected (SR) group (i.e., children who, from one grade level to the next, cycled in and out of rejection two or more times across the kindergarten to Grade 6 school years; $n = 17$); (e) stable chronic rejected (SCR) group (i.e., children rejected in five or more grades across the kindergarten to Grade 6 school years; $n = 51$); and (f) never rejected (NR) group (i.e., children who were never rejected across the kindergarten to Grade 6 school years; $n = 178$).

The validity of these a priori rejection groups was examined by also forming trajectory groups from participant's averaged peer acceptance/rejection ratings using growth mixture modeling (GMM; Muthén & Muthén, 1998–2004). Results showed that a four-trajectory model best fit the data (e.g., four-class model: AIC = 2,299.37, BIC = 2,342.71; three-class model: AIC = 6,067.05, BIC = 6,137.97; two-class model: AIC = 6,071.09, BIC = 6,130.20) and that these empirically derived trajectory classes included groups that exhibited (a) a declining peer acceptance trajectory (decliners), (b) an increasing peer acceptance trajectory (gainers), (c) a high-level peer acceptance trajectory (stable nonrejectees), and (d) a low-level peer acceptance trajectory (stable rejectees). Cross-tabulation of participants who were assigned to the a priori versus GMM trajectory groups indicated that except for the 1-YR and SR groups, there was substantial overlap in group membership. That is, 76% of the participants assigned to the LCR group were identified by GMM as decliners, 89% of children in the ECR group were classified as gainers, 86% of participants in the SCR group were designated stable rejectees, and 96% of participants in the NR group were identified as stable nonrejectees. Most likely, GMM did not identify trajectory classes corresponding to the 1-YR or SR groups because the participants in these a priori groups exhibited highly heterogeneous trajectories (e.g., these individuals moved in or out of rejection at different time points during the kindergarten to Grade 6 school years).

Third, analyses were undertaken to determine whether the six groups differed from each other with respect to participant's sex, race, aggressive and withdrawn behavior, and family SEI. Separate one-factor analyses of variance (ANOVAs; groups) were calculated on the three continuous variates (i.e., the behavioral and SEI scores), and z tests for proportions were computed for the two categorical variates (i.e., sex and race).

Group differences ($p < .05$) were found for sex, race, and family SEI. No significant group differences were found for either withdrawn or aggressive behavior. Sex composition differed in that whereas boys outnumbered girls by approximately 2:1 in all of the groups comprising children who were rejected more than 1 year (i.e., the ECR, LCR, SCR, and SR groups), girls outnumbered boys by approximately 2:1 in groups comprising children rejected 1 year or less (i.e., the 1-YR and NR groups). This sex difference was consistent with previous research, which has shown that more boys than girls tend to be rejected within their peer groups (Coie, Dodge, & Coppotelli, 1982; O'Neil, Welsh, Parke, Wang, & Strand, 1997). When the proportions of males within the ECR, LCR, SCR, and SR groups were compared, no significant differences were found, and the same was true for females within these groups ($z < 0.47$, *ns*). Racial composition did not differ in five of the six groups (LCR, SCR, SR, NR, and 1-YR) and averaged 79.6% European American, 16.9% African American, 1.2% Latino/Hispanic, 2.9% mixed race, and 0.2% Native American across these groups. In contrast to these groups, the ECR group contained a larger proportion of African Americans

(37.8% vs. an average of 16.9%). SEI scores, averaged by group, ranged from 35.02 to 55.68. Children in the NR group ($M = 55.57$) had higher scores than those in the ECR ($M = 40.90$) and SCR ($M = 35.02$) groups, and children in the LCR group ($M = 55.68$) had higher scores than those in the SCR group ($M = 35.02$).

Analyses of the Identified Groups' Classroom Participation Trajectories

Premises about the relation between chronic peer rejection and children's classroom participation trajectories were evaluated in a series of growth curve analyses (Mplus; Muthén & Muthén, 1998–2004). These analyses were conducted to determine whether each of the six groups' observed participation trajectories conformed to expectation. The predicted trajectory patterns were as follows. For the ECR group, it was expected that movement out of chronic rejection (rejection cessation) at Grade 3 would be accompanied by either no change in participation (i.e., trajectory continuity—continuation of the preexisting stable or downward trend) or a shift toward growth in participation (i.e., trajectory discontinuity) from Grades 3 to 6. For the LCR group, a shift from stable or positive growth to negative or declining growth in participation was expected as children transitioned into chronic rejection at Grade 3. For the SCR group, it was expected that stable chronic rejection would be associated with a sustained lack of participation (i.e., stable nonparticipation) or progressively decreasing classroom participation from kindergarten to Grade 6. For the 1-YR and SR groups, significant changes in participation trajectories were not anticipated because children's exposure(s) to peer rejection were transitive (i.e., brief or sporadic) rather than chronic. Finally, for the NR group, no significant changes in participation trajectories were expected because these children were not exposed to peer rejection during kindergarten to Grade 6.

To determine whether the data were consistent with these hypotheses, growth curve models were computed to estimate initial group differences in participation (intercept or between-group differences) and within-group growth rates in participation across kindergarten to Grade 6 (both linear and curvilinear trends). Additionally, for the two groups in which significant trajectory discontinuities were expected (i.e., the ECR and LCR groups), group-specific piecewise growth curves (Singer & Willett, 2003) also were estimated. Thus, in addition to initial status (intercepts), these group's growth rates (slopes) were estimated twice—in one case, before and after the transition from chronic rejection to acceptance (for the ECR group), and in the other case, before and after the transition from acceptance to rejection (for the LCR group). Because between-group differences were observed in sex, ethnicity, and family SES, these measures were included as covariates in all analyses.

For each analysis, unconditional models (Level 1) were initially calculated to estimate the mean trajectory for all children in the sample and to determine whether variation in individuals' trajectories (i.e., intercepts and linear slopes) relative to the mean trajectory on the targeted participation variable was sufficiently large to warrant the computation of conditional models. Quadratic and cubic trends were also tested in these unconditional models but were dropped because the variation associated with these growth parameters was statistically not significant. Conditional models (Level 2) were calculated because results for the unconditional models indicated that significant variation was present in the intercepts and/or linear slopes of individuals' trajectories ($ps > .05$). In these analyses, participation trajectories for the six groups were estimated by effects coding the categorical "groups" variable (Curran, Bauer, & Willoughby, 2004).

Independent participation. The results for this analysis are shown in Table 1 and are graphically represented in Figure 1a. Intercepts or initial status for all groups differed significantly from zero ($ts > 19.81$, $ps < .001$), and a one-factor ANOVA (groups: ECR, LCR, SCR, 1-YR, SR, and NR) performed on these means showed that although the initial (kindergarten) independent participation scores for the ECR and SCR groups did not differ significantly from each other (*ns*), both were significantly lower than those for the NR, 1-YR, SR, and LCR groups ($p < .05$). Tests on slopes indicated the following: For participants in the ECR group, the trajectory for independent participation remained flat or stable during the grades that they were rejected (kindergarten to Grade 3: −0.04, $t = -0.84$, *ns*) but increased significantly thereafter, during the grades that they were no longer rejected (Grades 4–6: 0.07, $t = 2.08$, $p < .05$). In contrast, for children in the LCR group, independent participation increased significantly across kindergarten to Grade 3 (0.05, $t = 2.35$, $p < .05$, when not rejected) but declined significantly across Grades 4–6 after they became rejected (−0.02, $t = -1.98$, $p < .05$). Participants in the SCR group had an independent participation trajectory that remained low and flat (stable) throughout grade school (−0.02, $t = -0.50$, *ns*). Equivalent slopes were found for the 1-YR, SR, and NR groups, and all were indicative of significant positive growth in independent participation from kindergarten to Grade 6 (0.03, $t = 3.83$, $p < .001$).

Cooperative participation. Results for this conditional growth model are shown in Table 1 and are

Table 1

Results of Piecewise Growth Curve Analysis for Chronically Rejected and Nonrejected Groups: Intercept and Slope Estimates for Independent and Cooperative Participation

	Independent participation			Cooperative participation		
Level of model	Coefficient	*SE*	*t*	Coefficient	*SE*	*t*
Models for initial status						
NR, SR, and 1-YR	2.61	0.03	84.77***	2.79	0.02	130.96***
SCR	1.79	0.08	21.97***	2.13	0.06	33.78***
LCR	2.46	0.12	19.81***	2.64	0.08	31.39***
ECR	2.00	0.13	19.94***	2.29	0.08	27.20***
Models for linear change						
Slope for NR, SR, and 1-YR	0.03	0.01	3.83***	0.01	0.01	0.81
Slope for SCR	−0.02	0.03	−0.50***	0.06	0.04	1.78
Slope prior to rejection, LCR	0.05	0.03	2.35***	0.02	0.02	0.86
Slope during rejection, LCR	−0.02	0.01	−1.98***	−0.02	0.01	−1.96*
Slope during rejection, ECR	−0.04	0.04	−0.84**	0.03	0.03	1.08*
Slope after cessation of rejection, ECR	0.07	0.03	2.08***	0.04	0.02	1.97*

Note. 1-YR = 1-year rejected; ECR = early chronic rejected; LCR = later chronic rejected; NR = never rejected; SR = sporadically rejected.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

graphically depicted in Figure 1b. Intercepts or initial status for all groups differed significantly from zero (*t*s > 27.20, *p*s < .001). An ANOVA (groups: ECR, LCR, SCR, 1-YR, SR and NR) performed on children's cooperative participation scores at kindergarten showed that the differences between the ECR and SCR groups and between the LCR, 1-YR, SR, and NR groups were not significant (*ns*). Estimates of linear change showed the following: For participants in the ECR group, the trajectory for cooperative participation remained flat or stable during the grades that they were rejected (kindergarten to Grade 3: 0.03, $t = 1.08$, *ns*) but increased significantly thereafter, during the grades that they were no longer rejected (Grades 4–6: 0.04, $t = 1.97$, $p < .05$). The cooperative participation trajectory for children in the LCR group, in contrast, remained moderately high and stable during the grades that they were not rejected (kindergarten to Grade 3: 0.02, $t = 0.86$, *ns*) but descended significantly during subsequent grades when they were rejected (Grades 4–6: −0.02, $t = -1.96$, $p < .05$). Participants in the SCR group had a cooperative participation trajectory that remained low and flat (stable) throughout grade school (0.06, $t = 1.78$, *ns*). For the 1-YR, SR, and NR groups, cooperative participation remained high and stable across kindergarten to Grade 6 (0.01, $t = 0.81$, *ns*).

Discussion

The results of this investigation serve to clarify how classroom peer rejection is related to children's school engagement and, specifically, how children's movement in and out of chronic peer rejection is associated with changes in their independent and cooperative participation trajectories during the grade school years. In large part, this investigation's merits derive from the fact that its findings reflect on several fundamental but previously underresearched and untested hypotheses.

Our first aim was to evaluate the hypothesis that chronic peer rejection is a detriment to classroom engagement and, in particular, a feature of classroom social environments that inhibits children's independent and cooperative classroom participation. This hypothesis was corroborated by evidence indicating that although children were chronically rejected, they exhibited minimal or no growth in classroom participation. This relation was substantiated within each of the groups that experienced periods of chronic rejection. Slope estimates showed that children's participation trajectories were flat or declining during the periods of grade school that they were chronically rejected—specifically, the initial half of grade school for the ECR group, the latter half of grade school for the LCR group, and all the grade school for the SCR group.

It is significant that regardless of whether chronic rejection was experienced by children during the early or later years of grade school, its presence was accompanied by insubstantial growth in classroom participation. Children who were chronically rejected only during the beginning of their school careers—that is, from kindergarten to Grade 3 (i.e., the ECR

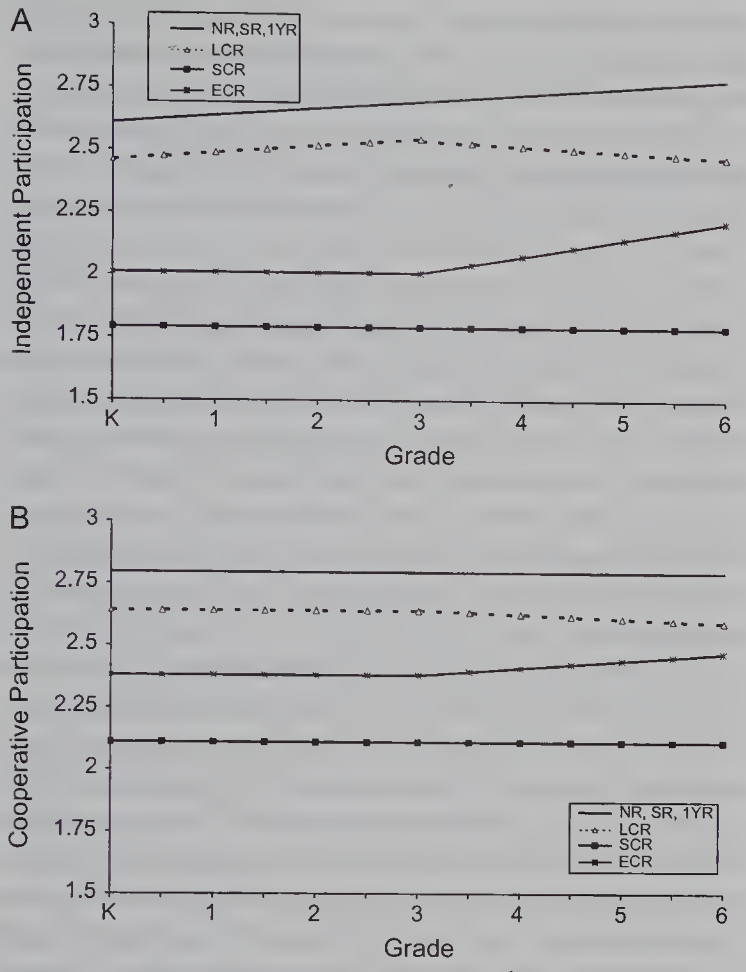


Figure 1. (a) Piecewise growth curve results for chronic rejection groups: Group-specific discontinuity model for independent participation showing differential shifts in slope when transitioning between rejection and nonrejection. (b) Piecewise growth curve results for chronic rejection groups: Group-specific discontinuity model for cooperative participation showing differential shifts in slope when transitioning between rejection and nonrejection.

group)—exhibited a significantly lower level of participation relative to nonrejected and transiently rejected children and little or no growth in participation throughout this period. This low and stable trajectory not only corroborated the premise that chronic peer rejection inhibits children's classroom participation but also was consistent with the logic that chronic rejection's effects accrue quickly and persist over time. For the children in the ECR group, chronic rejection was related to a substantial participation deficit that was evident soon after they began grade school and that persisted throughout the primary grades. Likewise, children who were not chronically rejected during kindergarten to Grade 3 but became so during Grades 4–6 (i.e., those in the LCR group) exhibited a significant downward inflection in their participation trajectories that began soon after they transitioned into chronic peer rejection. Prior to this point (i.e., during the primary grades), the participation trajectories for children in this group closely resembled those of nonrejected or transiently rejected children (i.e., children in the NR, SR, and 1-YR groups). That is, children in the LCR group exhibited significantly higher levels of participation than did children who were chronically rejected during this period (i.e., their counterparts in the ECR and SCR groups), and they showed significant positive growth in independent participation and stable levels of cooperative participation across grades kindergarten to Grade 3.

Furthermore, in all the groups that experienced chronic rather than sporadic or transient rejection (i.e., ECR, LCR, and SCR groups), the participation slopes that were estimated concurrently with chronic rejection showed that rejection was accompanied by little or no growth in *either* form of classroom participation. If rejection's role in this relation is causal (an inference that is plausible but not provable given this study's design), then these findings suggest that its "effects" on children's classroom participation were consistent with the hypothesis that chronic peer rejection is a detriment to children's independent *and* cooperative classroom participation.

Also examined as part of this first aim was the hypothesis that following a transition away from rejection and toward acceptance among classmates, children's participation trajectories would either (a) endure or worsen over time (continuity hypothesis) or (b) improve over time (discontinuity hypothesis). The evidence obtained for the ECR group provided the best test of this hypothesis because these children experienced a sustained period of peer rejection (from kindergarten to Grade 3) that was followed by a nearly equal interval of nonrejection (from Grades 4 to 6). Growth estimates for this group showed that the relatively flat participation trajectory that accompanied chronic rejection from kindergarten to Grade 3 was followed by a positive turnabout in both forms of classroom participation. That is, in contrast to the participation slopes estimated during chronic rejection (i.e., for kindergarten to Grade 3), the slopes estimated after rejection ceased (i.e., for Grades 4–6) marked the presence of significant growth in children's independent and cooperative classroom participation.

Rather than suggesting that peer rejection orients children toward enduring or irreversible participation trajectories, these findings imply that the transition from rejection toward acceptance is accompanied by an increase in classroom participation. Thus, the obtained data patterns contradicted the continuity hypothesis, which holds that peer rejection orients children toward enduring, irreversible disengagement trajectories—ones that are maintained even *after* they are no longer rejected by their classmates. Rather, the finding lent support to the discontinuity hypothesis or the view that the cessation of rejection permits growth in classroom participation and, therefore, deflects children from their preexisting participation trajectories. These findings not only corroborated the discontinuity hypothesis but also lent support to the inference that the cessation of *chronic* rejection frees children from enduring interpersonal constraints that may inhibit their classroom participation over substantial periods of time. If this inference is valid, then our data would suggest that once children are free of these constraints, they become progressively more engaged (cooperatively and independently) within classrooms.

Our data also reflected on the hypothesis that longer periods of chronic rejection inhibit classroom participation to a greater extent than shorter periods of chronic peer rejection. To corroborate this hypothesis, it would have been necessary to find that longer periods of chronic rejection (i.e., as exhibited by children in the SCR vs. those in the ECR or the LCR groups) were associated with (a) more enduring periods of stability or decline (negligible or negative growth) and/or (b) steeper rates of decline (more rapidly decelerating growth) in children's participation trajectories. As is illustrated in the following paragraphs, support was found for the former but not for the latter of these two possibilities.

Results indicated that children in the SCR group (i.e., those rejected a minimum of 5 years and a maximum of 7 years) manifested lower levels of classroom participation over a more protracted period of time compared to children in the other two chronic rejection groups (i.e., the ECR or LCR groups in which children were rejected a maximum of 2–3 years). Thus, children whose exposure to peer rejection lasted longest not only began school with low levels of both forms of participation but also exhibited no significant growth in either form of engagement throughout the entire grade school period. Moreover, this was the case regardless of the form of participation examined. That is, children in the SCR group exhibited the most sustained patterns of disengagement for both independent and cooperative participation. Essentially, children in the SCR group were relatively disengaged throughout grade school, whereas those in the other two chronic rejection groups—that is, the ECR and LCR groups—remained disengaged only as long as they continued to be rejected by their classmates. These findings further substantiate the premise that peer rejection inhibits children's classroom participation.

The evidence obtained here did not provide definitive support for the proposition that longer periods of chronic rejection produce more rapid rates of descent in children's classroom participation. Although children in the LCR group did exhibit trajectory discontinuities that were consistent with this hypothesized pattern (i.e., growth followed by significant decline in participation as children moved into chronic rejection), the participation trajectories of the other two chronic rejection groups were not characterized by change (declines) over time. Rather, stability characterized the cooperative participation trajectories of children in the ECR and SCR groups. As previously noted, for these children, chronic rejection corresponded to a substantial participation deficit that was evident during their 1st year of grade school and that was maintained as long as they continued to be rejected. It may be the case that these children's participation levels were reduced to such low levels as they began grade school that there was little room for further decline (e.g., floor effects occurred).

If it is the case that chronic peer rejection takes a negative toll on children's classroom participation, then it will be important to examine the processes that account for this relation. Potential targets for investigation include many of the intervening variables that were stipulated in the conceptual framework and hypotheses that guided this investigation. These include the many types of peer behaviors perpetrated against chronically rejected children that impede or discourage classroom participation (e.g., ignoring, exclusion, and harassment) and the plethora of rejection-induced psychological states that may produce similar consequences (e.g., fear of interacting with peers, reduced interest or motivation to participate in school activities, self-imposed isolation to escape peers' rejecting behaviors, and debilitating emotional states such as learned helplessness, anxiety, loneliness). Most, if not all, of these factors represent plausible avenues through which rejection may interfere with or reduce children's school engagement, and several have already been implicated empirically as barriers to classroom engagement, learning, and achievement (Buhs & Ladd, 2001; Fredricks, Blumenfeld, & Paris, 2004; Ladd, 1989, 2003, 2005).

Also worthy of consideration are emergent hypotheses and findings that implicate neuropsychological processes as potential mediators between interpersonal rejection and children's school performance. Rejection as an interpersonal experience has been linked with neurological activity in brain sites that are associated with pain and discomfort (Eisenberger, Lieberman, & Williams, 2003). There is also evidence to suggest that brain activity of this type impairs executive functions and thus may serve to dysregulate children's ability to focus, concentrate, and learn in school situations (Vailancourt, Clinton, McDougall, Schmidt, & Hymel, in press). Thus, it is possible that chronic peer rejection impedes children's school performance, including their ability to engage and profit from classroom activities because it dysregulates or alters brain functions that are central to learning and memory.

With respect to this investigation's second aim, the results did not provide strong support for the hypothesis that chronic peer rejection during grade school interferes with independent participation more than it does cooperative participation. Instead, the findings were more congruent with the premise that chronic rejection has similar rather than different effects on the targeted forms of classroom participation. The evidence indicated that both independent and cooperative participation changed little while children were chronically rejected, and the fact that these trajectories were fairly static raises the possibility that rejection delays rather than reduces each of these aspects of classroom engagement. The only evidence that implied that these two forms of participation develop differently was found in growth rates that were observed for nonrejected and transiently rejected children. The growth patterns for children in these groups (i.e., NR, SR, and 1-YR groups) were characterized by positive growth for independent participation and stability for cooperative participation. Furthermore, as previously noted, this same pattern typified children in the LCR group during the period of grade school when they were not chronically rejected by peers (i.e., kindergarten to Grade 3). Thus, unlike children who were chronically rejected during kindergarten to Grade 3, those who were nonrejected or transiently rejected exhibited greater growth in independent than in cooperative participation. These findings implied that there were differences in the way these two forms of participation developed but only for children who were not chronically rejected by peers. Results such as these suggest that normatively—that is, in the absence of persistent peer rejection—children tend to show stronger growth in independent participation compared to cooperative participation across the grade school period. Those who are exposed to chronic rejection, in contrast, appear to be at risk for delays or reductions in both forms of engagement.

Clearly, caution is warranted in the interpretation of these findings because even though relevant covariates were examined, other unmeasured factors (e.g., "third" variables) could have co-occurred with the documented patterns of peer group rejection. Consequently, further investigation is needed to evaluate the premises tested here, and innovative research designs and controls will be required to achieve stronger causal inferences. However, if the present findings are valid, they raise an intriguing implication: It may be possible for children to "recover" from the effects of chronic peer rejection—that is, become more engaged in classroom activities after they are no longer rejected by their classmates. These findings can be seen as a positive precedent within a body of research evidence that generally depicts few, if any, favorable outcomes for peer-rejected children.

References

Altman, J. (1974). Observational study of behavior: Sampling methods. *Behavior, 49*, 227–267.

Asher, S. R., & Dodge, K. A. (1986). Identifying children who are rejected by their peers. *Developmental Psychology, 22*, 442–449.

Asher, S. R., Rose, A. J., & Gabriel, S. W. (2001). Peer rejection in everyday life. In M. R. Leary (Ed.), *Interpersonal rejection* (pp. 105–142). London: Oxford University Press.

Asher, S. R., Singleton, L. C., Tinsley, B. R., & Hymel, S. (1979). A reliable sociometric measure for preschool children. *Developmental Psychology, 15*, 443–444.

Birch, S. H., & Ladd, G. W. (1997). The teacher-child relationship and children's early school adjustment. *Journal of School Psychology, 35*, 61–79.

Blumenfeld, P. C., Hamilton, V. L., Bossert, S. T., Wessels, K., & Meece, J. (1983). Teacher talk and student thought: Socialization into the student role. In J. M. Levine & M. C. Wang (Eds.), *Teacher and student perceptions: Implications for learning* (pp. 143–192). Hillsdale, NJ: Lawrence Erlbaum.

Buhs, E. S., & Ladd, G. W. (2001). Peer rejection in kindergarten as an antecedent of young children's school adjustment: An examination of mediating processes. *Developmental Psychology, 37*, 550–560.

Buhs, E. S., Ladd, G. W., & Herald, S. L. (2006). Peer exclusion and peer abuse: Processes that mediate the relation between peer group rejection and children's classroom engagement and achievement? *Journal of Educational Psychology, 98*, 1–13.

Bukowski, W. M., & Hoza, B. (1989). Popularity and friendship: Issues in theory, measurement, and outcome. In T. J. Berndt & G. W. Ladd (Eds.), *Peer relationships in child development* (pp. 15–45). New York: Wiley.

Burks, V. S., Dodge, K. A., & Price. J. M. (1995). Models of internalizing outcomes of early rejection. *Development and Psychopathology, 7*, 683–695.

Cillessen, A. H. N., & Bukowski, W. M. (Eds.). (2000). Recent advances in the measurement of acceptance and rejection in the peer system. *New Directions for Child Development* (No. 88). San Francisco: Jossey-Bass.

Coie, J. D. (1990). Towards a theory of peer rejection. In S. R. Asher & J. D. Coie (Eds.), *Peer rejection in childhood* (pp. 365–401). New York: Cambridge University Press.

Coie, J. D., Dodge, K. A., & Coppotelli, H. (1982). Dimensions and types of social status: A cross-age perspective. *Developmental Psychology, 18*, 557–570.

Curran, P. J., Bauer, D. J., & Willoughby, M. T. (2004). Testing main effects and interactions in latent curve analysis. *Psychological Methods, 9*, 220–237.

Damon, W., & Phelps, E. (1989). Strategic uses of peer learning in children's education. In T. J. Berndt & G. W. Ladd (Eds.), *Peer relationships in child development* (pp. 135–157). New York: John Wiley & Sons.

DeRosier, M. E., Kupersmidt, J. B., & Patterson, C. J. (1994). Children's academic and behavioral adjustment as a function of the chronicity and proximity of peer rejection. *Child Development, 65*, 1799–1813.

Dodge, K. (1983). Behavioral antecedents of peer status. *Child Development, 54*, 1386–1399.

Dohrenwend, B. S., & Dohrenwend, B. P. (1981). Life stress and illness: Formulations of the issues. In B. S. Dohrenwend & B. P. Dohrenwend (Eds.), *Stressful life events and their contexts* (Vol. 1, pp. 1–27). New York: Prodist.

Eisenberger, N. I., Lieberman, M. D., & Williams, K. D. (2003). Does rejection hurt? An fMRI study of social exclusion. *Science, 302*, 290–292.

Entwisle, D. R., & Astone, N. M. (1994). Some practical guidelines for measuring youth's race/ethnicity and sociometric status. *Child Development, 65*, 1521–1540.

Finn, J. D. (1989). Withdrawing from school. *Review of Educational Research, 59*, 117–142.

Finn, J. D. (1993). *School engagement and students at risk*. Washington, DC: Department of Education, National Center for Educational Statistics (ERIC Document Reproduction Service No. ED 362 322).

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*, 59–109.

Furman, W., & Gavin, L. A. (1989). Peer's influence on adjustment and development: A view from the intervention literature. In T. J. Berndt & G. W. Ladd (Eds.), *Peer relationships in child development* (pp. 319–340). New York: Wiley.

Johnson, J. H. (1988). *Life events as stressors in childhood and adolescence*. Newbury Park, CA: Sage.

Kochenderfer, B. J., & Ladd, G. W. (1996). Peer victimization: Cause or consequence of school maladjustment? *Child Development, 67*, 1305–1317.

Kochenderfer-Ladd, B., & Wardrop, J. (2001). Chronicity and instability in children's peer victimization experiences as predictors of loneliness and social satisfaction trajectories. *Child Development, 72*, 134–151.

Ladd, G. W. (1989). Toward a further understanding of peer relationships and their contributors to child development. In T. J. Berndt & G. W. Ladd (Eds.), *Peer relationships in child development* (pp. 1–11). New York: Wiley.

Ladd, G. W. (1990). Having friends, keeping friends, making friends, and being liked by peers in the classroom: Predictors of children's early school adjustment? *Child Development, 61*, 1081–1100.

Ladd, G. W. (2003). Probing the adaptive significance of children's behavior and relationships in the school context: A child by environment perspective. In R. Kail (Ed.), *Advances in child behavior and development* (pp. 43–104). New York: Wiley.

Ladd, G. W. (2005). *Children's peer relations and social competence: A century of progress*. New Haven, CT: Yale University Press.

Ladd, G. W., Birch, S. H., & Buhs, E. S. (1999). Children's social and scholastic lives in kindergarten: Related spheres of influence? *Child Development, 70*, 1373–1400.

Ladd, G. W., Buhs, E. S., & Seid, M. (2000). Children's initial sentiments about kindergarten: Is school liking an antecedent of early classroom participation and achievement? *Merrill-Palmer Quarterly, 46*, 255–279.

Ladd, G. W., Buhs, E. S., & Troop, W. (2002). Children's interpersonal skills and relationships in school settings: Adaptive significance and implications for school-based prevention and intervention programs. In P. K. Smith & C. H. Hart (Eds.), *Blackwell's handbook of childhood social development* (pp. 394–415). London: Blackwell.

Ladd, G. W., & Burgess, K. B. (2001). Do relational risks and protective factors moderate the linkages between childhood aggression and early psychological and school adjustment? *Child Development, 72*, 1579–1601.

Ladd, G. W., Herald, S., & Andrews, K. (2006). Young children's peer relations and social competence. In B. Spodek & O. Saracho (Eds.), Handbook of research on the education of young children (Vol. 2, pp. 23–54). New York: Macmillan.

Ladd, G. W., Kochenderfer, B. J., & Coleman, C. C. (1996). Friendship quality as a predictor of young children's early school adjustment. *Child Development, 67*, 1103–1118.

Ladd, G. W., Kochenderfer, B. J., & Coleman, C. C. (1997). Classroom peer acceptance, friendship, and victimization: Distinct relational systems that contribute uniquely to children's school adjustment? *Child Development, 68*, 1181–1197.

Ladd, G. W., Price, J. M., & Hart, C. H. (1990). *Preschoolers' behavioral orientations and patterns of peer contact: Predictive of social status?* In S. R. Asher & J. D. Coie (Eds.), *Peer rejection in childhood* (pp. 90–118). New York: Cambridge University Press.

Ladd, G. W., & Profilet, S. M. (1996). The child behavior scale: A teacher-report measure of young children's aggressive, withdrawn, and prosocial behaviors. *Developmental Psychology, 32*, 1008–1024.

Ladd, G. W., & Troop-Gordon, W. (2003). The role of chronic peer difficulties in the development of children's psychological adjustment problems. *Child Development, 74*, 1344–1367.

MacDougall, P., Hymel, S., Vaillancourt, T., & Mercer, L. (2001). The consequences of childhood peer rejection. In M. R. Leary (Ed.), *Interpersonal rejection* (pp. 213–247). Oxford, England: Oxford University Press.

Muthén, L. K., & Muthén, B. O. (1998–2004). *Mplus user's guide* (3rd ed.). Los Angeles, CA: Muthén & Muthén.

O'Neil, R., Welsh, M., Parke, R. D., Wang, S., & Strand, C. (1997). A longitudinal assessment of the academic correlates of early peer acceptance and rejection. *Journal of Clinical Child Psychology, 26*, 290–303.

Parker, J. G., & Asher, S. R. (1987). Peer relations and later personal adjustment: Are low-accepted children at risk? *Psychological Bulletin, 102*, 357–389.

Parker, J. G., Rubin, K. H., Price, J. M., & DeRosier, M. (1995). Peer relationships, child development, and adjustment: A developmental psychopathology perspective. In D. Cicchetti & D. Cohen (Eds.), *Developmental psychopathology: Vol. 2. Risk, disorder and adaptation* (pp. 96–161). New York: Wiley.

Perry, K. E., & Weinstein, R. S. (1998). The social context of early schooling and children's school adjustment. *Educational Psychologist, 33*, 177–194.

Putallaz, M., & Gottman, J. M. (1981). An interactional model of children's entry into peer groups. *Child Development, 52*, 986–994.

Radziszewska, B., & Rogoff, B. (1988). Influence of adult and peer collaboration on the development of children's planning skills. *Developmental Psychology, 24*, 840–848.

Rubin, D. B. (1987). *Multiple imputation for non-response in surveys*. New York: Wiley.

Schafer, J. L. (1999). NORM. Retrieved April 28, 2006, from http://www.stat.psu.edu

Singer, J. D. & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. London: Oxford University Press.

Slavin, R. (1983). When does cooperative learning increase student achievement? *Psychological Bulletin, 94*, 429–445.

Vandell, D. L., & Hembree, S. E. (1994). Peer social status and friendship: Independent contributors to children's social and academic adjustment. *Merrill-Palmer Quarterly, 40*, 461–470.

Vailancourt, T., Clinton, J., McDougall, P., Schmidt, L., & Hymel, S. (in press). The neurobiology of peer victimization and rejection. In S. R. Jimerson, S. M. Swearer, & D. L. Espelage (Eds.), *The International handbook of school bullying*. Mahwah, NJ: Erlbaum.

Wentzel, K. R. (1991). Social competence at school: Relation between social responsibility and academic achievement. *Review of Educational Research, 61*, 1–24.

Child Development, July/August 2008, Volume 79, Number 4, Pages 1016–1031

Numerical Magnitude Representations Influence Arithmetic Learning

Julie L. Booth and Robert S. Siegler
Carnegie Mellon University

This study examined whether the quality of first graders' (mean age = 7.2 years) numerical magnitude representations is correlated with, predictive of, and causally related to their arithmetic learning. The children's pretest numerical magnitude representations were found to be correlated with their pretest arithmetic knowledge and to be predictive of their learning of answers to unfamiliar arithmetic problems. The relation to learning of unfamiliar problems remained after controlling for prior arithmetic knowledge, short-term memory for numbers, and math achievement test scores. Moreover, presenting randomly chosen children with accurate visual representations of the magnitudes of addends and sums improved their learning of the answers to the problems. Thus, representations of numerical magnitude are both correlationally and causally related to arithmetic learning.

Recent studies have revealed a consistent and robust age-related change in children's representations of numerical magnitudes. With age and experience, children progress on many tasks from relying on logarithmic representations of numerical magnitudes to relying on linear representations (Booth & Siegler, 2006; Geary, Hoard, Byrd-Craven, Nugent, & Numtee, 2007; Opfer, Thompson, & DeVries, 2007; Siegler & Booth, 2004; Siegler & Opfer, 2003). The meaning of this finding is illustrated in Figure 1, which displays representative data on number line estimation. When children are presented number lines with endpoints of 0 and 100 and no marks or numbers in between and then are asked to estimate the positions of specific numbers on the number line, most kindergartners produce estimates that closely fit a logarithmic function. In contrast, most second graders generate a linear pattern of estimates; about half of first graders produce each of the two patterns. As can be seen in the figure, use of a logarithmic representation entails estimates increasing more rapidly for smaller numbers than for larger ones, whereas use of a linear representation entails estimates increasing at an equal rate throughout the numerical range.

This shift from logarithmic to linear patterns of number line estimates has been found to occur between kindergarten and second grade for 0–100 number lines and between second and fourth grades for 0–1,000 ones (Booth & Siegler, 2006; Geary et al., 2007; Laski & Siegler, 2007; Siegler & Booth, 2004). The same shift from logarithmic to linear estimation patterns occurs at the same ages on other estimation tasks, in particular, measurement estimation (drawing a line of approximately *N* units, where line lengths of 0 and 1,000 units are present as reference points) and numerosity estimation (generating approximately *N* dots on a computer screen, where 1 and 1,000 dots are the reference points; Booth & Siegler, 2006). A parallel change has been found for kindergartners', first graders', and second graders' categorization of the numbers 0–100 (Laski & Siegler, 2007).

The consistency of individual differences in performance on these tasks also attests to the effects of the changes in numerical magnitude representation. Within-grade correlations of individual second and fourth graders' degree of linearity on number line estimation, measurement estimation, and numerosity estimation tasks ranged from $r = .50$ to $r = .80$ in two experiments (Booth & Siegler, 2006). Within-grade correlations of individual kindergartners', first graders', and second graders' linearity of numerical categorization and linearity of number line estimation ranged from $r = .55$ to $r = .85$ in two other experiments (Laski & Siegler, 2007). Within-grade correlations between individual kindergartners' through fourth graders' math achievement test scores and their linearity of number line estimation have typically ranged from $r = .50$ to $r = .60$ (Booth & Siegler, 2006; Siegler & Booth, 2004).

Funding for this research was provided by the National Institutes of Health Grant HD 19011 and the Office of Educational Research and Instruction Grant R305H020060. Thanks are also due to the Hopewell Area School District for allowing us to collect data in their schools and to Mary Wolfson for assistance with data collection.

Correspondence concerning this article should be addressed to Julie L. Booth, Human Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, PA 15213, or to Robert S. Siegler, Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213. Electronic mail may be sent to juliebooth@cmu.edu or rs7k@andrew.cmu.edu.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7904-0014

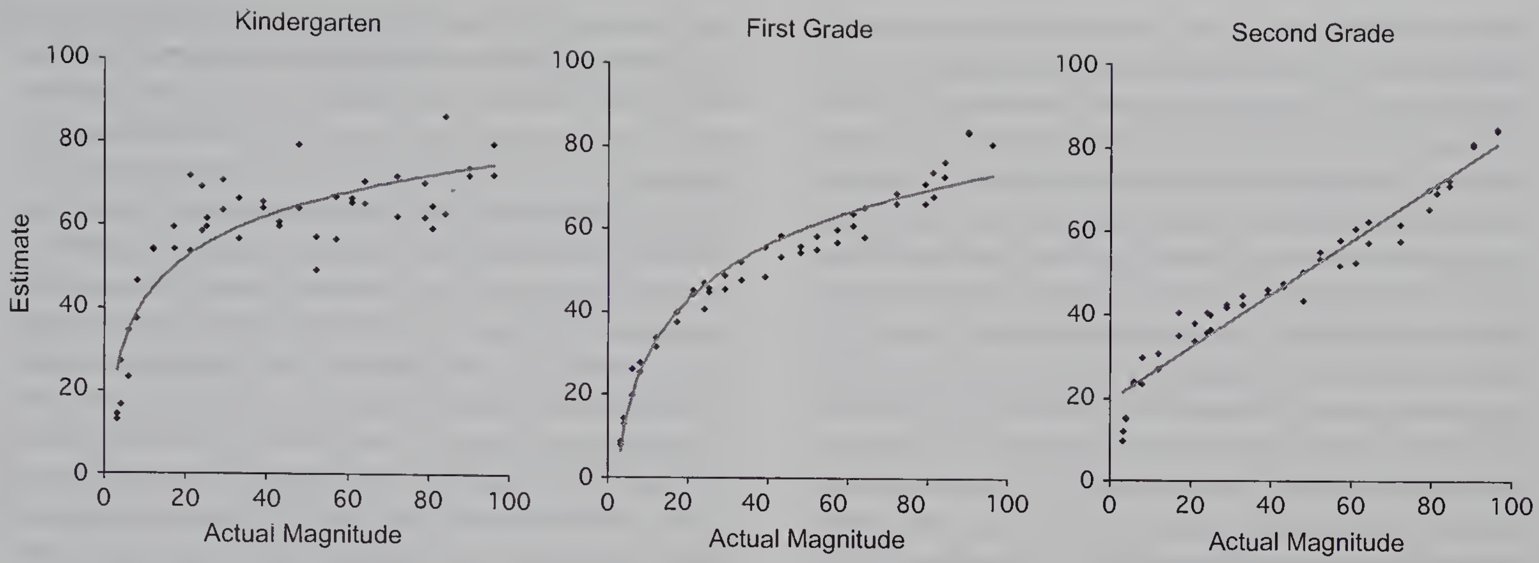


Figure 1. Median number line estimates and best-fitting function in Siegler and Booth (2004).

The central question of the present study was whether linearity of numerical magnitude representations influences acquisition of *new* numerical information. This was the first study to address this question. The particular numerical acquisition that was examined was learning answers to unfamiliar arithmetic problems. Though learning arithmetic facts has generally been viewed as a rote memorization process (e.g., Ashcraft, 1992; Campbell & Graham, 1985), we believe that such learning also includes meaningful cognitive activity. In particular, it seems to involve relying not only on verbal associations between problems and answers but also on number sense and an understanding of how quantities are combined to generate answers with particular magnitudes (see Jordan, Kaplan, Locuniak, & Ramineni, 2007, and Baroody, 1985, for similar views). Thus, there was reason to believe that children's representations of numerical magnitudes might influence their arithmetic learning.

Studies of the effects of preschool mathematics curricula provided another reason to hypothesize that numerical magnitude representations influence learning of answers to arithmetic problems. Several preschool programs, in particular, the Right Start and Number Worlds curricula (Griffin, 2004; Griffin, Case, & Siegler, 1994), are designed to inculcate a mental number line representation, akin to the linear representation of numerical magnitudes hypothesized in the present research. Griffin, Case, and their colleagues (Griffin, 2004; Griffin et al., 1994) found that having low-income preschoolers participate in a wide range of numerical activities that were intended to stimulate such a representation—numerical board games, computer games, monetary activities, number-related projects, number-related songs, and so on—improved the children's mathematical performance in first grade.

It seems likely that the extensive experience with numbers provided by Project Right Start and Number Worlds contributes to well-differentiated knowledge of a wide range of numerical magnitudes (i.e., knowledge in which students discriminate clearly among the magnitudes of large as well as small numbers rather than just among the magnitudes of small numbers). Well-differentiated knowledge of both large and small numbers seems likely to constrain the range of potential answers that receive more than minimal activation when arithmetic problems are presented. Connectionist models posit that multiple soft constraints limit potential responses to any given problem by suppressing activations of unlikely responses and boosting activation of more likely ones (e.g., McClelland, Rumelhart, & Hinton, 1986). This leads to most errors being close misses as well as to increasingly accurate performance, both of which characterize learning of arithmetic (Geary, 2006).

The implications of this analysis for arithmetic learning may be clarified by imagining a first grader being asked whether 24 + 18 might equal 82. A child who used a linear representation of the magnitudes of numbers between 0 and 100 would quickly answer "no" because the problem would elicit little if any activation of such a large answer. In contrast, for a child with a logarithmic representation of numerical magnitudes between 0 and 100, 82 would not be such an implausible answer because the logarithmic representations of 42 and 82 are not hugely different (3.74 and 4.41). Perhaps for that reason, children who generated logarithmic patterns of number line estimates were more likely to categorize 42 and 82 as being within the same numerical category than were children who generated linear estimation patterns (Laski & Siegler, 2007). Thus, linear representations of numerical magnitudes should increase the likelihood that errors on relatively large arithmetic problems will be close misses and should also increase the likelihood of retrieving correct answers.

Errors and solution time patterns on verification tasks are consistent with this analysis and with the interpretation that representations of numerical magnitudes influence arithmetic performance above and beyond knowledge of the correct answers. Both children and adults more quickly reject errors that are far from the correct sum than errors whose magnitudes are closer to the correct value (Ashcraft, Fierman, & Bartolotta, 1984; Campbell & Graham, 1985). For this and the previously described reasons, we hypothesized that the linearity of representations of numerical magnitudes would influence learning of unfamiliar arithmetic problems.

The Current Study

The present study had three main purposes. One was to examine whether first graders' representations of the magnitudes of numbers between 0 and 100 are related to their existing arithmetic knowledge and to their ability to acquire new knowledge of arithmetic problems in this numerical range. For the reasons described in the previous section, both relations seemed plausible. First graders seemed the ideal age group for testing this hypothesis because, at least within the United States, this is the grade in which there are the most substantial individual differences in representations of numerical magnitudes between 0 and 100 (Laski & Siegler, 2007; Siegler & Booth, 2004).

A second purpose of the study was to examine the relation to arithmetic learning of several other variables that seemed likely to influence such learning: overall mathematical knowledge (as indexed by scores on standardized achievement tests), prior knowledge of addition, and short-term memory for numbers. Although many studies of arithmetic knowledge have been conducted, we were unable to locate any that systematically examined relations of prior numerical knowledge to acquisition of new arithmetic knowledge under experimentally controlled conditions that guaranteed equal exposure to the problems. The present study was designed to provide information on this basic question.

In addition to being of interest in itself, examination of these predictive relations also made possible a rigorous test of the hypothesis that numerical magnitude representations influence arithmetic learning above and beyond general mathematical proficiency. If the linearity of numerical magnitude representations per se is influential, then measures of the linearity of the magnitude representations should predict learning of unfamiliar arithmetic problems above and beyond the influences of standardized math achievement test scores, prior knowledge of addition, and short-term memory for numbers.

The third major purpose of the study was to test whether provision of visual representations of numerical magnitudes exerts a causal influence on learning of answers to addition problems. External representations, including pictorial, graphic, and diagrammatic forms, are an important part of mathematics education (Seeger, 1998). They are thought to increase understanding of mathematical concepts by helping children build relations among mathematical ideas (Hiebert & Carpenter, 1992). Consistent with this claim, the National Council of Teachers of Mathematics (NCTM) recommends that teachers include multiple forms of representations when teaching mathematical concepts (NCTM, 2000). Evidence supporting this recommendation is rather weak, however. For example, Sowell (1989) concluded that the use of physical manipulatives is beneficial, but Beishuizen (1993) found that their usefulness may be limited to children with low levels of achievement. Indeed, in some contexts, provision of manipulatives appears to be harmful rather than helpful to mathematics learning (Uttal, Liu, & DeLoache, 2006). Similarly, findings from some studies suggest that provision of pictorial information is beneficial to math learning (e.g., Beishuizen, 1993; Klein, Beishuizen, & Treffers, 1998), whereas findings from others indicate that many students do not benefit from them (e.g., Cobb, 1995; Gravemeijer, 1994). Which types of external representations are most helpful and when such representations are helpful are poorly understood.

To address these issues, the first graders who participated in the present study were randomly assigned to encounter one of several types of external representations of the magnitudes of addends and sums as they attempted to learn unfamiliar addition facts. Children in the *computer generate* group were presented 0 – 100 number lines with computer-generated (correct) representations of the magnitudes of addends and sums for four unfamiliar addition problems as they tried to learn the answers to the problems. Peers in the *child generate* group were asked to estimate for themselves on 0 – 100 number lines the magnitudes of the addends and the sums as they tried to learn the answers to the four trained problems. Children in the *child and computer generate* condition first generated magnitude representations for themselves and then saw the computer-generated (correct) representations as they tried to learn the answers to the problems. Finally, classmates in the *control* group were given identical exposure to the problems and answers but were neither asked to generate a visual representation of them nor presented the accurate magnitude representations.

This design allowed us to test four main hypotheses about arithmetic learning. The first hypothesis was that linearity of numerical magnitude representations on the pretest predicts arithmetic learning. The logic underlying this prediction was that linear magnitude representations would lead children to activate a relatively narrow range of potential answers that included the correct sum rather than activating either an overly broad range of numbers that did not provide much constraint on plausible answers or a range that did not include the sum. In contrast, children who relied on a logarithmic representation of the numbers between 0 and 100 would activate a broad range of numbers in the numerical range into which answers to the four trained problems fell (sums of 27–92) because the logarithmic representation does not discriminate strongly among them. (Within a linear representation, the magnitudes of 92 and 27 differ by a ratio of 3.41:1, whereas in a logarithmic representation, the magnitudes differ by a ratio of only 1.37:1.) Depending on how the representations of the addends were implemented in the context of solving arithmetic problems, the activated range might not even include the correct answer. Consider 17 + 29, one of the problems on which children received training. The natural logs of the addends sum to 6.20, whereas the natural log of 46 is 3.83. If children use the natural log values of 17 and 29 to generate an approximate answer, the correct sum might be activated minimally if at all.

The second hypothesis was that the predictive relation between pretest linearity of numerical representations and arithmetic learning would be present above and beyond the relations to arithmetic learning of math achievement test scores, prior knowledge of other addition problems, and short-term memory for numbers. All of these other variables were expected to be related to arithmetic learning, but the relation of the linearity of each child's numerical magnitude representation to that child's arithmetic learning was expected to be present even after the other variables were considered. This prediction was a particularly rigorous test of the hypothesized relation between numerical magnitude representations and arithmetic learning because math achievement test scores share substantial common variance with the linearity of numerical magnitude representations (Booth & Siegler, 2006; Siegler & Booth, 2004). Moreover, math achievement tests have far better psychometric properties than the 5-min assessment of children's numerical magnitude representations that we used. Nonetheless, because math achievement and the other variables were hypothesized to reflect not only magnitude representations but also a variety of other influences and because magnitude representations were hypothesized to be particularly important for learning new arithmetic problems, linearity of numerical representations (as measured by linearity of number line estimates) was expected to explain variance in arithmetic learning beyond that which could be accounted for by the other three variables.

The third hypothesis was that children who were given accurate information about numerical magnitudes (those in the *computer generate* and *child and computer generate* groups) would show more learning of the correct sums, and greater improvements in the distance of their answers from the correct sum, than children in the two groups that did not receive such information. Encountering accurate visual representations of the numerical magnitudes of the addends and sum of each problem would help children align the combined magnitudes of the addends with the magnitude of the corresponding sum. It also would provide a visual processing route to the approximate value of the sum, which would support the verbal association of problem and answer.

The fourth hypothesis was that children who generated their own estimates of the magnitudes of the addends and sum (those in the *child generate* and *child and computer generate* conditions) would show greater improvement in knowledge of the trained addition problems than children who did not generate visual representations of the magnitudes of addends and sums for themselves. Asking children to generate representations of addends and sums was expected to lead to greater engagement in the task and deeper processing of the numerical magnitudes. Experimental manipulations that improve depth of processing have been found to enhance learning on a variety of mathematical and scientific problems, including addition (Calin-Jageman & Ratner, 2005), balance scales (Pine & Messer, 2000), and mathematical equality (McNeill, 2007). A similar effect was anticipated in the present context.

Method

Participants

The children who participated, 105 first graders in the second half of their academic year (57 boys and 48 girls, mean age = 7.20 years, *SD* = 0.40), were recruited from four public elementary schools in a predominately middle-income school district. The percentages of children at the schools who were eligible for free or reduced-cost lunches were 14%, 15%, 22%, and 24%, respectively. Almost all participants (96%) were Caucasian, as were both experimenters.

Children were randomly assigned to one of four conditions: *child generate* ($N = 26$), *computer generate* ($N = 25$), *child and computer generate* ($N = 27$), or *control* ($N = 27$). The slightly unequal sample sizes in the four groups were due to two children being absent when the 2-week follow-up session was conducted and one child not completing the posttest number line task due to computer malfunction. Participation was completely voluntary and no extrinsic rewards were provided.

Tasks

Number Line Estimation

Knowledge of numerical magnitudes was assessed using a computerized number line estimation task. On each problem, at the top of the screen was a number between 1 and 99 and in the middle of the screen was a horizontal number line with 0 at the left end and 100 at the right end. The experimenter explained that the number line always went from 0 to 100 and that the child should click on the line to show where he or she thought the number at the top should go. Children were first given two practice trials, which involved marking 100 and then 0 on the number line. If children did not place each practice number at the appropriate endpoint, they were shown where the number belonged. After this orienting task, the remaining numbers were presented, one at a time without feedback.

To facilitate our ability to discriminate between linear and logarithmic estimation patterns, we slightly oversampled the numbers at the low end of the 0 – 100 range by including four numbers from each decade below 30 and two numbers from each successive decade. The 26 numbers used on the number line task were 2, 3, 6, 7, 11, 14, 15, 19, 21, 23, 24, 28, 32, 36, 44, 47, 51, 58, 63, 69, 72, 76, 84, 87, 91, and 98. Children received one of three random orderings of the numbers in each session, with the orderings counterbalanced across sessions.

Math Achievement Test

The math section of the Wide Range Achievement Test – Expanded (WRAT – Expanded) was administered to obtain a measure of each child's general math knowledge. This standardized test includes questions on counting, reading numerals, measurement, spatial reasoning, and oral and written addition and subtraction. It was developed based on the curriculum content guidelines and standards developed by the NCTM (2000) and has been shown to be reliable in terms of internal consistency and stability (Robertson, 2001).

Short-Term Memory Test

Short-term memory for numbers was measured using the standard forward digit span task from the Wechsler Intelligence Scale for Children – Fourth Edition (Wechsler, 2002). The experimenter told children that she would say some numbers and that the child's job was to repeat the numbers after she stopped. Children were given one practice trial with the string "1, 2" and were corrected if they did not provide the correct response. On all trials, the presentation rate was 1 per second. The experimental task began with two trials of three digits each. If the child succeeded on at least one trial, the length of the list was increased by one digit after two presentations of that length. If the child failed on both presentations at a given level, the test was terminated. The child's score was the value of the longest series of digits correctly repeated.

Addition

Thirteen addition problems with answers between 0 and 100 were used to examine the first graders' knowledge of addition: 26 + 27, 1 + 4, 18 + 16, 38 + 39, 5 + 12, 49 + 43, 13 + 9, 48 + 17, 8 + 5, 17 + 29, 5 + 4, 46 + 35, and 9 + 18. With the exception of the few problems with very small sums, these problems were likely to be unfamiliar to first graders. All the problems with sums greater than 20 involved carrying; such problems were chosen to prevent children from solving these problems via strategies other than retrieval and thus to require them to use retrieval. A time limit of 6 s per problem was imposed for the same reason.

The instructions also conveyed the information that children should not count or use algorithms to solve the problems but rather should try to retrieve the answer. The experimenter told children that their job was to say a number that they thought could be the answer, rather than trying to add the numbers in their head, and that all of the answers would be between 0 and 100. The experimenter presented each problem verbally and asked the child to give the answer quickly. On the rare occasions when a child did not answer quickly (i.e., within 6 s after the experimenter finished reading the problem), the experimenter reminded the child not to calculate the answer, just to say a number that could be the answer if they did not know it. All children received the 13 addition problems in the same random order.

Four of the 13 addition problems were used as training items. Pilot testing indicated that these four problems were ones that the first graders rarely answered correctly on the pretest. The problems included one item that tended to elicit answers that were relatively close to the correct answer (9 + 18), two that elicited answers that tended to be incorrect by a moderate amount (26 + 27 and 17 + 29), and one that tended to elicit answers that were highly discrepant from the correct answer (49 + 43). In each session, after stating their answers for all of the problems, children in all conditions were told that they would be shown the answers to the problems and that they should try really hard to learn the correct answers. During each training session, students encountered four presentations of the first problem, followed by four presentations of the second problem, then the third, and finally the fourth. At the end of each training session, children were given a no-feedback test of the four trained addition facts, in which the experimenter presented each problem verbally and asked the child to give the answer.

Instructional Procedures

During the three training sessions, the particular procedures and types of visual information encountered by children varied with their instructional condition. Children in the *computer generate* group were shown horizontally oriented 0–100 number lines with red, blue, and purple bars that correctly represented the magnitudes of the addends and sum. Peers in the *child generate* group generated bars on number lines for themselves to estimate the magnitudes of the addends and sums; therefore, their visual representations could be accurate or inaccurate. Children in the *child and computer generate* group received both visual representations. Finally, children in the *control* group were presented the same problems and answers on the screen but without any visual representation of the numerical magnitudes. Each instructional condition is described in more detail subsequently.

Computer Generate Condition

Children were told that they would see the problem and the answer on the computer screen, that their job was to watch as the computer indicated the sizes of the numbers on the number line, and to try really hard to learn the answers to the problems. They were then given a practice problem to orient them to the important features of the task: that the red and blue bars on the computer screen represented how big the addends were and that the purple bar for the answer equaled the length of the bars for the two addends laid end to end.

The practice problem had "10 + 10" at the top of the computer screen and a line with 0 and 100 at the two ends in the middle of the screen. The experimenter went over this problem twice in somewhat different ways. First, the experimenter read the problem aloud and asked the child to repeat it. The first 10 was then highlighted in red and circled on the screen, and the child was asked to read the red number. Next, the correct magnitude for 10 was indicated in a three-step procedure: First, a small vertical hatch mark appeared on the number line; then, a red bar from 0 to the hatch mark appeared; and then, 10 was displayed above the center of the red bar. The experimenter said, "The red line shows how big 10 is. Notice that the line is red, just like the number in the problem. The first number is always going to be red." The red bar and the number above it then disappeared, and the procedure was repeated for the second 10 (which was represented with a circled blue number and a blue line). Finally, the sum, which was represented with a circled purple number and a purple bar, was presented.

At the beginning of the second time through the practice problem, the computer screen was reset to its original state. This time, the child read the first 10, which was again highlighted in red, and the screen displayed its magnitude with a red bar. The experimenter reminded the child that the red bar showed how big the first addend was. Then, the child read the second 10, and the screen displayed its magnitude with a blue bar under the number line, with the first addend and the red bar representing its magnitude remaining visible. The experimenter reminded the child that the blue bar indicated how big the second addend was and said, "Let's add the red line for 10 to the blue line for 10." At this point, the blue bar, which represented the second addend, was moved to the top of the number line, beginning where the red bar ended. Then, the purple bar that represented the magnitude of the answer appeared just under the number line, and the experimenter explained that the purple bar indicated how big 20 was and that the child should notice that the length of the red bar plus the length of the blue bar equaled the length of the purple bar. Finally, with this display remaining visible, the child was asked to read the problem one final time.

After the practice problem, children were tested on their knowledge of what the red, blue, and purple bars meant. The experimenter asked the same three questions until the child answered all of them correctly. Then, the four training problems were presented along with the correct visual representation of each one's addends and sum.

Child Generate Condition

Children in this group were given the same instructions except that they were told that they (rather than the computer) would estimate the magnitudes of the numbers on the number line. The remainder of the procedure was identical to that in the *computer generate* condition, except for the children estimating the magnitudes of the addends and sum on the number line. Red, blue, and purple bars corresponding to the child's estimates of the addends and sum for each problem (instead of the computer-generated estimates) appeared on the screen as in the *computer generate* condition.

Child and Computer Generate Condition

Children in this group were given the same initial instructions and procedure as peers in the *child generate* group. However, the second time through the problem, the screen also showed computer-generated (correct) placements on a separate number line that was positioned directly underneath the child-generated estimates for the problem. For example, for the problem 17 + 29, after the child's estimate of the magnitude of 17 appeared just above the upper number line, the computer generated the correct magnitude of 17 just above the lower number line. The experimenter then said, "This red line (pointing) is how big you thought 17 was, but *actually* this red line (pointing) is how big 17 is." Then, after the child's estimate of the magnitude of 29 appeared in blue just below the upper number line, the correct magnitude of 29 appeared in blue just below the lower number line. Next, on both number lines, the blue bar representing the magnitude of the second addend was placed at the right end of the red bar that represented the magnitude of the first addend. After that, a purple bar representing the sum of the two addends appeared below the relevant number line. On the upper number line, the purple bar corresponded to the child's estimates of the sum, whereas on the lower number line, the purple bar corresponded to the correct sum. Thus, by the end of the trial, the computer screen displayed all three of the child's estimates on the upper number line and all of the correct magnitudes on the lower line.

Control Condition

As in the other three conditions, children in the *control* group encountered each problem four times in each training session. The conditions under which children were presented each problem were also the same as in the other conditions. First, the experimenter read the addends and sum aloud. Then, the child repeated them. The next two times through the problem, the addends and sum were colored red, blue, and purple and circled as in the other conditions, and the child read them without the experimenter's help.

Procedure

Children met one-on-one with the experimenter for three 10- to 15-min sessions within a 1-week period and for a 5-min "follow-up" session 2 weeks after the third session. All sessions were conducted during the school day; children participated at a time selected by their teacher. An outline of the tasks presented in each of the four sessions is presented in Table 1.

Table 1
Measures Included in Each of the Four Experimental Sessions

Measure	Session 1 (pretest)	Session 2	Session 3 (end of training)	Session 4 (follow-up)
Number line estimation	X		X	
Math achievement test		X		
Short-term memory	X			
13-item addition test	X			X
4-item addition pretest, training, and posttest procedure	X[a]	X	X	

[a] In Session 1, the 4-item addition pretest reflected performance on the 4 items on which children subsequently received training from within the 13-item addition test.

In Session 1, children were presented the number line estimation task, the short-term memory task, and the 13-problem addition task. Then, they were presented the instructional procedure for their experimental group, followed by a no-feedback test on the four trained addition problems at the end of the session. All children received the tasks in the same order. Session 2 consisted of a pretest, training, and a posttest on the four addition training problems, followed by the math section of the WRAT. Session 3 was identical to Session 2, except that children were given the number line estimation task rather than the WRAT at the end of the session. During Session 4 (the follow-up session), children were again asked to provide sums for the same 13 addition problems as on the pretest.

No feedback regarding answers to the trained problems was given on the pretests and posttests for those problems, and only implicit feedback (i.e., the correct answer) was given during training. Children were told frequently during all parts of the experiment that they were doing well. At the end of each session, they were thanked for participating and returned to their classroom.

Results

The results are reported in two sections. In the first section, we examine for the entire sample the interrelations of performance on the four pretest tasks—addition, number line estimation, short-term memory for numbers, and math achievement test—and examine each of their predictive relation to learning of the four trained addition problems. In the second section, we examine the causal influence of computer-generated and self-generated visual representations on the learning of the trained arithmetic problems and on number line estimates. The analyses examined performance at three times: *pretest* (Session 1), *end of training* (end of Session 3), and *follow-up* (Session 4).

Predictors of Learning and Performance

Relations Among Pretest Measures

As shown in Table 2, levels of performance on all four pretest measures of numerical proficiency were positively related. The relations among number line estimation, addition, and math achievement test scores were substantial; the relations between short-term memory and the other three tasks were weaker. The correlation between linearity of number line estimates and overall math achievement test score (r = .57) was of the same magnitude as in previous studies (rs = .50 – .60).

Table 2
Correlations Among Number Line Estimation, Addition, Short-Term Memory, and Math Achievement Test Performance: Pretest Data

Measure	Number line estimation (R^2_{Lin})	Math achievement (WRAT)	Addition (PAE)
Math achievement (WRAT)	.57**		
Addition (PAE)	.41**	.48**	
Short-term memory	.07	.21*	.21*

Note. df = 103. PAE = percent absolute error; WRAT = Wide Range Achievement Test.
$*p < .05$. $**p < .01$.

Pretest Predictors of Arithmetic Learning

We examined pretest predictors of performance on the trained addition problems at the end of training and at the follow-up session for the whole sample and for each instructional condition separately. The predictive relations were similar for all conditions; therefore, only the results for the full sample are reported.

Addition knowledge. Children answered 21% of pretest items correctly. The large majority of correct answers on the pretest were on the four problems that had sums below 20—55%, versus 6% on the nine larger problems. On the four large addend problems on which children subsequently received training, they answered correctly on 7% of trials.

Knowledge of addition was also assessed through examining percent absolute error (PAE):

$$\left(\left|\frac{\text{Child's answer} - \text{correct answer}}{\text{Scale of answers}}\right|\right)$$

For example, if a child answered 80 for the problem 26 + 27, the PAE would be 0.27 ([80 – 53]/100) or 27%. PAE provides a continuous measure of the degree to which answers are of the right magnitude, which was a central issue in the present study.

Mean PAE on the pretest was 12%. Again, there was a large discrepancy between performance on the four small addend problems (PAE = 2%) and on the nine large addend ones (PAE = 17%). On the four training items, pretest PAE was 16%.

We next examined the stability of individual differences in addition knowledge, in particular, whether children's pretest addition knowledge was predictive of their posttest knowledge of the trained addition problems. To avoid the autocorrelations that would have been entailed by including the same items as predictors and predicted variables, we correlated each child's mean pretest PAE on the nine nontrained addition problems with that child's performance following training on the four trained addition problems. (Too few children answered any of these four problems correctly on the pretest for pretest number correct to be a meaningful predictor.)

PAE on the nine untrained problems on the pretest was positively related to PAE on the four trained addition problems at the end of training, $r(103) = .47$, $p < .01$, and at the follow-up, $r(101) = .62, p < .01$. PAE on the nine untrained problems on the pretest also was correlated with percent correct answers on the four trained problems both at the end of training, $r(103) = -.23, p < .05$, and at the follow-up, $r(101) = -.37, p < .01$.

We next conducted a regression analysis to determine whether pretest PAE on the nine untrained addition problems accounted for variance in learning of the four training set items beyond that which could be explained by pretest PAE on the training set items themselves. In this analysis, we first entered each child's pretest PAE on the four training set items and then their pretest PAE on the nine untrained addition problems.

As shown in Table 3, pretest PAE on the nine untrained addition problems added significant variance both at the end of training (8%) and at the follow-up (11%) beyond that which could be explained by pretest PAE on the trained problems. This finding suggests that children's general sense of the likely magnitudes of answers to addition problems predicts their subsequent learning, beyond their initial knowledge of answers to the specific problems that they are learning.

Number line estimation. To assess the relation between pretest numerical magnitude representations and addition learning, we correlated each child's R^2_{Lin} on the pretest number line estimation task with that child's percent correct answers and PAE on the trained addition facts at the end of training and at the follow-up.

R^2_{Lin} on the number line pretest predicted PAE on the trained addition facts both at the end of training, $r(103) = -.45, p < .01$, and at the follow-up, $r(101) = -.54, p < .01$. Pretest R^2_{Lin} on the number line pretest also predicted percent correct answers on the trained addition problems at the end of training, $r(103) = .23$, $p < .05$, and at the follow-up, $r(101) = .30, p < .01$. Thus, initial knowledge of numerical magnitudes predicted both learning of unfamiliar addition facts and quality of errors on them.

To determine whether pretest number line estimation was related to improvement in PAE on the addition facts, above and beyond the quality of initial answers on the problems, we conducted a regression analysis in which we first entered each child's pretest PAE on the trained addition facts and then their R^2_{Lin} values on the number line pretest. As shown in Table 3, R^2_{Lin} on the number line pretest added a substantial amount of variance to that which could be explained by pretest PAE on the trained addition

Table 3

Pretest Predictors of Improvement in PAE on Trained Addition Problems at End of Training and Follow-Up Performance

Pretest predictor(s)	% variance explained in reduction of PAE on trained problems at end of training	% variance explained in reduction of PAE on trained problems at follow-up
Trained addition problems (PAE)	31	18
Trained addition problems (PAE) + untrained addition problems (PAE)	39*	29*
Trained addition problems (PAE) + number line estimation (R^2_{Lin})	39*	29*
Trained addition problems (PAE) + WRAT score	41*	24*
Trained addition problems (PAE) + STM	33	20

Note. First row of data indicates variance in improvement in PAE on trained addition problems after training and at the follow-up that could be explained by pretest PAE on those problems. Subsequent rows indicate the variance in improvement in PAE at the end of training and at the follow-up on the trained addition problems that could be explained by considering pretest PAE on the same problems and the additional variable indicated in the left column. Thus, considering pretest number line estimation (R^2_{Lin}) as well as pretest PAE on the trained addition problems accounted for 8% more variance after training (39% vs. 31%) and 11% more variance at the follow-up (29% vs. 18%) than only considering pretest PAE on the same problems. PAE = percent absolute error; WRAT = Wide Range Achievement Test; STM = short-term memory.
*$p < .01$.

facts: 8% at the end of training and 11% at the 2-week follow-up.

Math achievement. Children's math achievement, as measured at the time of the pretest, also was predictive of their PAE on the trained addition problems both at the end of training, $r(103) = -.49, p < .01$, and at the follow-up, $r(101) = -.52$, $p < .01$. Their achievement test scores also were related to percent correct answers on the trained problems at the end of training, $r(103) = .28$, $p < .01$, and at the follow-up, $r(101) = .33$, $p < .01$. As shown in Table 3, the math achievement test scores, like pretest performance on the untrained addition problems and number line estimates, added to the variance that could be explained by pretest PAE on the trained addition facts: 10% additional variance at the end of training and 6% at the follow-up.

Short-term memory. Short-term memory for numbers was predictive of PAE on the trained addition facts both at the end of training, $r(103) = -.20, p < .05$, and at the follow-up, $r(101) = -.23, p < .05$. However, it was not related to percent correct on the trained addition problems either at the end of training, $r(103) = .11$, *ns*, or at the follow-up, $r(101) = .08$, *ns*. It also did not account for significant variance in improvement in PAE at the end of training or at the follow-up beyond that which could be explained by pretest PAE on the four trained addition problems (Table 3).

Linear representations as an independent predictor of addition learning. One of the main purposes of the study was to test the prediction that even after the effects of other aspects of numerical knowledge were considered, linearity of numerical magnitude representations would still predict arithmetic learning. Therefore, we conducted regression analyses on reduction of error on the trained addition facts by the end of training and at the follow-up. We first entered each child's pretest PAE on the trained addition facts, pretest PAE on untrained addition facts, pretest short-term memory for numbers, and math achievement test scores. Then, we entered the child's linearity of pretest number line estimates to see if it explained additional variance in the child's posttraining performance on the trained addition problems.

The linearity of pretest number line estimates added significant variance both at the end of training and at the follow-up to that which could be explained by the other variables. Pretest performance on the other four tasks explained 45% of the variance in PAE on the trained addition problems at the end of training; pretest R^2_{Lin} on the number line explained an additional 2%, $t(102) = 2.09, p < .05, d = 0.41$. The same four pretest variables explained 33% of the variance in PAE on the trained problems at the follow-up; pretest R^2_{Lin} on the number line estimation task explained an additional 6%, $t(100) = 2.88, p < .01$, $d = 0.58$. Thus, despite R^2_{Lin} on the number line task being correlated with three of the four other predictors and all of those other measures being correlated with PAE on the trained addition problems at the end of training and at the follow-up, linearity of number line estimates accounted for variance in arithmetic learning that could not be explained by the other measures of numerical knowledge.

Effects of Instructional Condition

To examine the effects on arithmetic learning of the experimental manipulations, we contrasted performance in the four experimental conditions. Comparisons were conducted on both number of correct answers and PAE.

Percent Correct on Trained Addition Problems

A 2 (computer-generated magnitude representations: present or absent) × 2 (child-generated magnitude representations: present or absent) × 2 (time of test: pretest or end of training) analysis of variance (ANOVA) on the number of correct answers was conducted to determine whether the visual representations of numerical magnitude influenced children's learning up to the end of training. The analysis yielded main effects for the time of test variable, $F(1, 101) = 70.14$, $p < .01$, $\eta^2 = .41$, and for exposure to computer-generated visual representations, $F(1, 101) = 4.08$, $p < .05$, $\eta^2 = .04$. Children generated more correct answers at the end of training than on the pretest (29% vs. 7%) and children who encountered the accurate, computer-generated representations of numerical magnitudes generated a greater number of correct answers than did children who did not encounter them (21% vs. 15%).

The interaction among the three variables also was significant, $F(1, 101) = 5.01$, $p < .05$, $\eta^2 = .05$. To interpret this interaction, we first conducted separate ANOVAs for pretest and end-of-training performance. No differences among the four groups were apparent on the pretest, $F(3, 101) = 1.02$, *ns*, $\eta^2 = .03$, but differences were present at the end of training, $F(3, 101) = 2.62$, $p = .05$, $\eta^2 = .07$. Fisher's least significant difference (LSD) post hoc tests indicated that at the end of training, children in the *computer generate* condition advanced a greater number of correct answers than peers in both the *control* and the *child and computer generate* conditions (*ps* < .05, *ds* = 0.75 and 0.61, respectively). These findings indicated that the interaction was due to the *computer generate* procedure increasing learning when presented alone but not when presented in combination with the *child generate* procedure.

Analyses of pretest–posttest changes for each of the four groups supported this interpretation. As shown in the top half of Table 4, children in all four groups answered correctly more often at the end of training than on the pretest: *control*, $t(26) = 3.43$, $p < .01$, $d = 0.88$; *child generate*, $t(25) = 4.41$, $p < .01$, $d = 1.22$; *computer generate*, $t(25) = 6.53$, $p < .01$, $d = 1.51$; and *child and computer generate*, $t(27) = 2.67$, $p < .01$, $d = 0.65$. However, the changes were larger in the *computer generate* condition (32% difference between pretest and posttest percent correct) than in the other three conditions (15%, 17%, and 24%).

A parallel 2 × 2 × 2 ANOVA on changes in correct answers between pretest and follow-up on the training problems indicated only a single main effect for time of testing, $F(1, 101) = 7.96$, $p < .01$, $\eta^2 = .07$. Children advanced a greater number of correct answers at the follow-up than at the pretest (13% vs. 7%).

PAE on Trained Addition Problems

To determine whether encountering accurate visual representations of the magnitudes of the addends and sum influenced the quality of students' errors as well as the frequency of correct answers, we conducted a 2 (computer-generated magnitude representations: present or absent) × 2 (child-generated magnitude representations: present or absent) × 2 (time of testing: pretest or end of training) ANOVA on PAE on the addition problems. As on the parallel analysis of correct answers, the ANOVA revealed a main effect for time of testing, $F(1, 101) = 6.59$, $p < .01$, $\eta^2 = .06$. From the pretest to the end of training, PAE on the addition problems dropped from 16% to 13%.

Table 4

Percent Correct and PAE for Trained Addition Problems on Pretest, End-of-Training, and Follow-Up Sessions

	Pretest	End of training	Follow-up
Percent correct			
Control	6 (13)	23 (24)	10 (16)
Child generate	4 (12)	28 (25)	13 (20)
Computer generate	9 (18)	41 (24)	13 (22)
Child and computer generate	10 (17)	25 (28)	14 (9)
PAE			
Control	18 (10)	16 (12)	18 (11)
Child generate	17 (12)	12 (9)	16 (13)
Computer generate	15 (11)	9 (9)	13 (10)
Child and computer generate	14 (9)	15 (10)	14 (9)

Note. All statistics are reported as *M* (*SD*). PAE = percent absolute error.

Also as on the parallel analysis of changes in correct answers, the ANOVA for the PAE data revealed an interaction among the three variables, $F(1, 101) = 4.36$, $p < .05$, $\eta^2 = .04$. The source of the interaction appeared to be the same as in the analysis of correct answers. Again, no differences among the four groups were detected on the pretest, $F(3, 101) = 0.62$, *ns*, $\eta^2 = .02$, but the groups tended to differ at the end of training, $F(3, 101) = 2.12$, $p = .10$, $\eta^2 = .06$. LSD post hoc tests indicated that children in the *computer generate* condition had a lower posttest PAE than peers in the *control* condition ($p < .05$, $d = 0.66$), and tended to have a lower posttest PAE than children in the *child and computer generate* condition ($p < .10$, $d = 0.63$).

Analyses of pretest–posttest changes for each of the four groups supported this interpretation (bottom half of Table 4). Children in the *computer generate* group reduced their PAE on the trained addition problems between the pretest and the end of training, $t(24) = 2.13$, $p < .05$, $d = 0.60$. There was also a trend toward reduced error for children in the *child generate* group between the pretest and the end of training, $t(25) = 1.88$, $p < .10$, $d = 0.47$. In contrast, PAE did not decrease either in the *child and computer generate* group, $t(26) = 0.32$, *ns*, $d = 0.11$, or in the *control* group, $t(26) = 0.84$, *ns*, $d = 0.18$.

A parallel 2 × 2 × 2 ANOVA on changes in PAE between pretest and follow-up on the training problems yielded no significant main effects or interactions.

Numerical Magnitude Estimation

Exposure to visual representations of numerical magnitudes was also expected to improve children's number line estimates because of the training providing experience with number line representations of numerical magnitudes. Separate 2 (computer-generated magnitude representations: present or absent) × 2 (child-generated magnitude representations: present or absent) × 2 (time of testing: pretest or end of training) ANOVAs were performed on PAE and linearity (R^2_{Lin}) of number line estimation.

Main effects of viewing computer-generated visual representations were found on PAE, $F(1, 100) = 7.23$, $p < .01$, $\eta^2 = .07$. The number line estimates of children who encountered accurate, computer-generated representations had a lower PAE than did peers who did not encounter the accurate visual representations (9% vs. 11%).

In addition to this main effect, an interaction between exposure to computer-generated representations and test session was present on the PAE measure, $F(1, 100) = 4.87$, $p < .05$, $\eta^2 = .05$. The number line estimates of children who viewed the computer-generated (correct) visual representations improved from a PAE of 10% on the pretest to 9% at the end of training. In contrast, the number line estimates of peers who did not see correct visual representations of numerical magnitude failed to improve: PAE increased from 11% to 12%.

Discussion

The results of this study supported three of the four hypotheses that motivated it. As hypothesized, the linearity of children's number line estimates correlated positively with their existing knowledge of addition. Also as hypothesized, the degree of linearity of children's pretest estimates was predictive of their learning of answers to novel addition problems. Moreover, providing accurate visual representations of the magnitudes of addends and sums increased children's learning of the novel addition problems beyond the level produced by simply presenting problems and answers. The one result that was contrary to expectation was that having children generate their own representations of addends and sums did not enhance learning and may have interfered with it. In this concluding section, we examine implications of these findings for understanding the role of numerical magnitude representations in childhood numerical competence, for understanding arithmetic learning, and for improving mathematics instruction.

Numerical Representations and Numerical Competence

The pretest results both replicated and extended previous findings that the linearity of elementary schoolchildren's representations of numerical magnitudes is related to their performance on a wide range of numerical tasks. Prior studies indicated that the linearity of number line estimates correlates positively with the linearity of measurement estimation, numerosity estimation, numerical categorization, and numerical magnitude comparison as well as with math achievement test scores (Booth & Siegler, 2006; Laski & Siegler, 2007; Siegler & Booth, 2004.) A very recent study added the findings that preschoolers' estimates on a 0–10 number line correlate positively and quite strongly with their counting, numeral identification, and magnitude comparison skills (Ramani & Siegler, 2008).

The present study replicated the relation of number line estimation to standardized achievement scores in those studies and indicated that first graders' linearity of numerical magnitude representations also is related to their arithmetic proficiency and short-term memory for numbers. Two other very recent studies have linked the linearity of number line estimation to yet other aspects of cognition: several measures of executive functioning and IQ test scores (Geary, Hoard, Nugent, & Byrd-Craven, in press; Geary et al., 2007). Geary and his colleagues advanced the intriguing hypothesis that the strong relations of executive functioning to linearity of number line estimates reflect the need for children to inhibit the previously dominant logarithmic representation when they begin to rely on the linear representation. Geary and colleagues' (Geary et al., 2007; Geary et al., in press) research also demonstrated that the number line estimates of first and second graders diagnosed as having math learning disabilities lag far behind those of typically achieving peers in both linearity and accuracy.

Taken together, the present and previous findings indicate that preschoolers' through fourth graders' numerical magnitude representations are positively correlated with performance on specific tasks measuring a wide range of specific numerical competencies as well as with performance on overall tests of mathematical achievement and IQ. The linearity of the numerical magnitude representations also shows robust and stable patterns of individual differences and consistent developmental trends (Booth & Siegler, 2006; Geary et al., 2007). Thus, as hypothesized by Case and colleagues (Case, 1992; Case & Okamoto, 1996; Griffin, Case, & Sandieson, 1992; Marini, 1992), the mental number line appears to provide a central conceptual structure for organizing a wide range of numerical knowledge.

Arithmetic Learning

The present findings indicated that numerical magnitude representations are not only positively related to a variety of types of numerical knowledge but also predictive of success in acquiring new numerical information, in particular, answers to arithmetic problems. Prior studies (e.g., Geary et al., 2007; Jordan et al., 2006) had shown that incoming characteristics of students predict their learning of arithmetic and other numerical skills during the school year. However, the present study appears to be the first to demonstrate that such predictive relations are present when learning occurs under experimentally controlled conditions.

An important implication of the present findings is that arithmetic learning, even in the sense of memorizing answers to unfamiliar problems, is not a rote activity but rather a meaningful one. If memorization of answers to arithmetic problems were a rote activity, there would be no reason to expect that being presented accurate visual representations of the magnitudes of addends and sums would increase learning of correct answers. Nor would there be any reason to expect that pretest linearity of number line estimation would predict learning of novel arithmetic problems, even after prior knowledge of arithmetic, short-term memory for numbers, and overall math achievement test scores were statistically controlled. These results cannot be explained in terms of exposure to the addends and answers, which were the same in all four conditions, nor in terms of time on task, which was greater for children generating their own representations of numerical magnitudes than for children who encountered the computer-generated accurate ones.

Several previous investigators have argued that backup strategies, such as counting fingers, counting from the minimum addend, and repeated addition, make arithmetic a meaningful activity (Brownell & Chazal, 1935; Cowan, 2003; Cowan, Dowker, Christakis, & Bailey, 1996; Cowan & Renton, 1996). The present findings add evidence that linear representations of numerical magnitudes make arithmetic meaningful even when children learn answers via retrieval.

Although the linearity of numerical magnitude representations on the pretest was related to improvements on the trained arithmetic problems in both percent correct and PAE, the relations to reductions in PAE were stronger, accounting for more than twice as much variance both at the end of training and at the follow-up. This finding was consistent with the theory underlying the experiment. The most straightforward way in which numerical magnitude representations would influence arithmetic learning is by constraining the set of possible answers to ones of approximately the right magnitude. Such constraints would have the direct effect of improving the quality of errors; they also would increase the likelihood of the exact correct answer because, by definition, it has exactly the right magnitude.

Linearity of numerical magnitude representations was not the only type of numerical knowledge that predicted arithmetic learning. Mathematical achievement test scores and addition knowledge also did. Yet, linearity of magnitude representations was predictive of arithmetic learning even after the relations of the other variables to arithmetic learning were statistically controlled. The question was why.

Our interpretation is that linearity of number line estimates varies more closely with the underlying variable of linearity of numerical magnitude representations than do the other two variables. Addition knowledge and achievement test scores vary in part with linearity of numerical magnitude representations, but they also vary with other influences (e.g., counting skills, memory for computational procedures, and short-term memory for numbers). Thus, linearity of number line estimation may predict learning of novel arithmetic problems above and beyond the prediction of the other variables because it provides a purer measure of the underlying magnitude representation.

The present findings ruled out the possibility that the predictive relation of linearity of number line estimation to arithmetic learning simply reflected general familiarity with numbers or the richness of the home intellectual environment. If either of these interpretations were correct, there would be no reason to expect that linearity of number line estimates would remain a significant predictor of learning of novel arithmetic problems even when achievement test scores and prior addition knowledge were statistically controlled. Indeed, there was reason to expect the opposite. Some parents teach their young children to solve arithmetic problems, whereas others do not. First graders' knowledge of addition would be expected to reflect this parental input. In contrast, to the best of our knowledge, no parents (not even the present authors) are so pedantic that they ask their children to locate numbers on number lines. Thus, the fact that the linearity of number line estimation predicts arithmetic learning above and beyond prior arithmetic knowledge indicates that learning of new arithmetic problems is influenced by numerical magnitude representations.

The one factor that, to our surprise, did not predict learning of addition facts was short-term memory for numbers. One possible explanation is that the training in this study allowed children to view the addends and the answers simultaneously as they were trying to learn them; thus, the task format may have reduced the burden on short-term memory of learning answers to arithmetic problems. Another possible explanation was that the digit span task was an inadequate measure of short-term memory capacity; a measure of working memory might have revealed stronger relations.

Instructional Implications

Results of the experimental manipulations yielded causal evidence that converged with the correlational data. As hypothesized, exposure to accurate visual representations of numerical magnitude improved children's arithmetic learning. This was true both when learning was measured in terms of percent correct and when it was measured in terms of PAE. Thus, the patterns of correlations on the pretest, the predictive relations from pretest performance to performance at the end of training, and the effects of providing randomly chosen children with accurate linear representations of numerical magnitude converged on the conclusion that linear representations of numerical magnitude promote arithmetic learning.

Self-generated representations of numerical magnitudes did not have a comparable effect. In fact, they appeared to interfere with the positive effects of being shown accurate visual displays when children received both. Thus, it appears that pictorial representations of numerical magnitudes must be accurate to enhance learning. The pattern of results dovetails with previous findings, indicating that being asked to explain correct answers improves learning of mathematical and scientific information, but being asked to explain the reasoning underlying ones' own answers, right or wrong, does not enhance learning (Siegler, 1995, 2002).

Perhaps the most surprising finding of the study was that the combination of computer-generated and self-generated magnitude representations was less effective in promoting learning than the computer-generated representations alone. One plausible interpretation was that the first graders in this condition simply could not take in all of the information when both representations were present. Seeing both visual representations on the screen may have provided a difficult source monitoring challenge in that it required remembering that the computer-generated representations were the accurate ones and that the self-generated ones were not necessarily accurate. Source monitoring is generally a difficult task for young children (Ratner, Foley, & Gimpert, 2002). The challenge might have been especially great in the current context because the correct interpretation required children to discount their own magnitude estimates.

The pattern of effects and noneffects in the present study highlights the value of presenting accurate visual representations of mathematical relations. Contemporary mathematics education for elementary schoolchildren makes a great deal of use of concrete objects that can be manipulated, such as Dienes Blocks and Cuisenaire Rods. Recent research, however, suggests that children often have difficulty viewing physical objects purely as symbols, even when that is the teacher's intent. This is especially the case when children handle or play with the objects (DeLoache & Burns, 1994; Uttal, Liu, & DeLoache, 2006).

The findings and analyses of DeLoache, Uttal, and their colleagues (DeLoache & Burns, 1994; Uttal et al., 2006), along with the present findings, provide a basis for hypothesizing that accurate pictorial representations, such as diagrams and graphs, may be particularly useful for promoting math learning. Like manipulatives, they provide a visual representation of the meaning of mathematical operations. Unlike manipulatives, they are not tempting objects for play and other activities that may interfere with their connection to mathematical operations. This analysis, together with the present finding that accurate visual representations of numerical magnitude promote arithmetic learning, suggests that providing pictorial information that illustrates mathematical relations in a transparent way can improve math learning as early as first grade.

References

Ashcraft, M. H. (1992). Cognitive arithmetic: A review of data and theory. *Cognition*, *44*, 75–106.

Ashcraft, M. H., Fierman, B. A., & Bartolotta, R. (1984). The production and verification tasks in mental addition: An empirical comparison. *Developmental Review*, *4*, 157–170.

Baroody, A. J. (1985). Mastery of basic number combinations: Internalization of relationships or facts? *Journal for Research in Mathematics Education*, *16*, 83–98.

Beishuizen, M. (1993). Mental strategies and materials or models for addition and subtraction up to 100 in Dutch second grades. *Journal for Research in Mathematics Education*, *24*, 294–323.

Booth, J. L., & Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology*, *41*, 189–201.

Brownell, W. A., & Chazal, C. B. (1935). The effects of premature drill on third grade arithmetic. *Journal of Educational Research*, *29*, 17–28.

Calin-Jageman, R. J., & Ratner, H. H. (2005). The role of encoding in the self-explanation effect. *Cognition and Instruction*, *23*, 523–543.

Campbell, J. I., & Graham, D. J. (1985). Mental multiplication skill: Structure, process, and acquisition. *Canadian Journal of Psychology Special Issue: Skill*, *39*, 338–366.

Case, R. (1992). *The mind's staircase: Exploring the conceptual underpinnings of children's thought and knowledge*. Hillsdale, NJ: Erlbaum.

Case, R., & Okamoto, Y. (1996). The role of central conceptual structures in the development of children's thought. *Monographs of the Society for Research in Child Development*, *61*(1–2).

Cobb, P. (1995). Cultural tools and mathematical learning: A case study. *Journal for Research in Mathematics Education*, *26*, 362–385.

Cowan, R. (2003). Does it all add up? Changes in children's knowledge of addition combinations, strategies, and principles. In A. J. Baroody & A. Dowker (Eds.), *The* *development of arithmetic concepts and skills: Constructing adaptive expertise* (pp. 35–74). Mahwah, NJ: Erlbaum.

Cowan, R., Dowker, A., Christakis, A., & Bailey, S. (1996). Even more precisely assessing children's understanding of the order-irrelevance principle. *Journal of Experimental Child Psychology, 62*, 84–101.

Cowan, R., & Renton, M. (1996). Do they know what they are doing? Children's use of economical addition strategies and knowledge of commutativity. *Educational Psychology, 16*, 407–420.

DeLoache, J. S., & Burns, N. M. (1994). Symbolic functioning in preschool children. *Journal of Applied Developmental Psychology, 15*, 513–527.

Geary, D. C. (2006). Development of mathematical understanding. In W. Damon & R. M. Lerner (Series Eds.) & D. Kuhn & R. S. Siegler (Vol. Eds.), *Handbook of child psychology: Vol. 2: Cognition, perception, and language* (6th ed., pp. 777–810). New York: Wiley.

Geary, D. C., Hoard, M. K., Byrd-Craven, J., Nugent, L., & Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Development, 78*, 1343–1359.

Geary, D. C., Hoard, M. K., Nugent, L., & Byrd-Craven, J. (in press). Development of number line representations in children with mathematical learning disability. *Developmental Neuropsychology.*

Gravemeijer, K. (1994). Educational development and developmental research in mathematics education. *Journal for Research in Mathematics Education, 25*, 443–471.

Griffin, S. (2004). Number worlds: A research-based mathematics program for young children. In D. H. Clements & J. Sarama (Eds.), *Engaging young children in mathematics: Standards for early mathematics education* (pp. 325–342). Mahwah, NJ: Erlbaum.

Griffin, S., Case, R., & Sandieson, R. (1992). Synchrony and asynchrony in the acquisition of children's everyday mathematical knowledge. In R. Case (Ed.), *The mind's staircase: Exploring the conceptual underpinnings of children's thought and knowledge* (pp. 75–98). Hillsdale, NJ: Erlbaum.

Griffin, S., Case, R., & Siegler, R. (1994). Rightstart: Providing the central conceptual prerequisites for first formal learning of arithmetic to students at risk for school failure. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 25–49). Cambridge, MA: MIT Press.

Hiebert, J., & Carpenter, T. P. (1992). Learning and teaching with understanding. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 65–97). New York: Macmillan.

Jordan, N. C., Kaplan, D., Locuniak, M. N., & Ramineni, C. (2007). Predicting first-grade math achievement from developmental number sense trajectories. *Learning Disabilities Research & Practice, 22*, 36–46.

Jordan, N. C., Kaplan, D., Olah, L. N., & Locuniak, M. N. (2006). Number sense growth in kindergarten: A longitudinal investigation of children at risk for mathematics difficulties. *Child Development, 77*, 153–175.

Klein, A. S., Beishuizen, M., & Treffers, A. (1998). The empty number line in Dutch second grades: Realistic versus gradual program design. *Journal for Research in Mathematics Education, 29*, 443–464.

Laski, E. V., & Siegler, R. S. (2007). Is 27 a big number? Correlational and causal connections among numerical categorization, number line estimation, and numerical magnitude comparison. *Child Development, 78*, 1723–1743.

Marini, Z. (1992). Synchrony and asynchrony in the development of children's scientific reasoning. In R. Case (Ed.), *The mind's staircase: Exploring the conceptual underpinnings of children's thought and knowledge* (pp. 55–74). Hillsdale, NJ: Erlbaum.

McClelland, J. L., Rumelhart, D. E., & Hinton, G. E. (1986). The appeal of parallel distributed processing. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing. Explorations in the microstructure of cognition: Vol. 1. Foundations* (pp. 3–44). Cambridge, MA: MIT Press.

McNeill, N. M. (2007, March). Tales of a fourth grade misconception: U-shaped development in children's performance on equivalence problems. In J. H. Bisanz (Chair), *Overcoming misconceptions: Mechanisms of positive change for a common mathematical misunderstanding*. Symposium conducted at the biennial meeting of the Society for Research in Child Development, Boston.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

Opfer, J. E., Thompson, C. A., & DeVries, J. (2007). Why children make better estimates of fractional magnitude than adults. In D. S. McNamara & G. Trafton (Eds.), *Proceedings of the 19th annual conference of the Cognitive Science Society.* Mahwah, NJ: Erlbaum.

Pine, K. J., & Messer, D. J. (2000). The effect of explaining another's actions on children's implicit theories of balance. *Cognition and Instruction, 18*, 35–51.

Ramani, G. B., & Siegler, R. S. (2008). Promoting broad and stable improvements in low-income children's numerical knowledge through playing number board games. *Child Development, 79*, 375–394.

Ratner, H. H., Foley, M. A., & Gimpert, N. (2002). The role of collaborative planning in children's source-monitoring errors and learning. *Journal of Experimental Psychology, 81*, 44–73.

Robertson, G. J. (2001). *Wide Range Achievement Test-Expanded*. Wilmington, DE: Wide Range.

Seeger, F. (1998). Representations in the mathematics classroom: Reflections and constructions. In F. Seeger, U. Waschescio, & J. Voight (Eds.), *The culture of the mathematics classroom* (pp. 308–343). New York: Cambridge University Press.

Siegler, R. S. (1995). How does change occur: A microgenetic study of number conservation. *Cognitive Psychology, 28*, 225–273.

Siegler, R. S. (2002). Microgenetic studies of self-explanations. In N. Granott & J. Parziale (Eds.), *Microdevelopment: Transition processes in development and learning* (pp. 31–58). New York: Cambridge University.

Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child Development*, *75*, 428–444.

Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science*, *14*, 237–243.

Sowell, E. J. (1989). Effect of manipulative materials in mathematics instruction. *Journal for Research in Mathematics Education*, *20*, 498–505.

Uttal, D. H., Liu, L. L., & DeLoache, J. S. (2006). Concreteness and symbolic development. In L. Balter & C. S. Tamis-LeMonda (Eds.), *Child psychology: A handbook of contemporary issues* (2nd ed., pp. 167–184). New York: Psychology Press.

Wechsler, D. (2002). *Wechsler Intelligence Scale for Children–IV.* San Antonio, TX: Psychological Corporation.

Child Development, July/August 2008, Volume 79, Number 4, Pages 1032–1048

Investigating the Development of Data Evaluation: The Role of Data Characteristics

Amy M. Masnick
Hofstra University

Bradley J. Morris
Grand Valley State University

A crucial skill in scientific and everyday reasoning is the ability to interpret data. The present study examined how data features influence data interpretation. In Experiment 1, one hundred and thirty-three 9-year-olds, 12-year-olds, and college students (mean age = 20 years) were shown a series of data sets that varied in the number of observations and the amount of variance between and within observations. Only limited context for the data was provided. In Experiment 2, similar data sets were presented to 101 participants from the same age groups incrementally rather than simultaneously. The results demonstrated that data characteristics affect how children interpret observations, with significant age-related increases in detecting multiple data characteristics, in using them in combination, and in explicit verbal descriptions of data interpretations.

One crucial skill in scientific and everyday reasoning is the ability to interpret data. For example, imagine two golfers on a driving range hit three balls each. Golfer A hits the balls 100, 130, and 125 yards, whereas Golfer B hits the balls 250, 265, and 270 yards. If asked to predict which golfer would likely hit the ball farther on the next drive, an observer would probably pick Golfer B. It is likely that our observer would not treat each drive as an unrelated incident but would see each drive as one piece of data related to a stable underlying construct (e.g., skill in driving a golf ball). Along with this understanding is an expectation that even when such a construct is stable, there is variation in outcomes (i.e., exact distance of drive). Although the mechanism causing the difference might not be clear to the observer (e.g., strength, practice), the resulting pattern of observations suggests nontrivial differences between the performances of the two golfers. Explanations for the difference in performance (i.e., specifying causal mechanism) would arise from a combination of the observer's background knowledge and the data.

Although such inferences may be trivial for adults, children often err when making similar comparisons (see Zimmerman, 2000, 2007, for reviews). One source of difficulty is that children tend to rely heavily on their domain knowledge when interpreting data, even to the extent of ignoring disconfirming data that conflict with their current knowledge or expectations (e.g., Chinn & Malhotra, 2002; Kuhn, Garcia-Mila, Zohar, & Andersen, 1995; Schauble, 1996). For example, when children observed the speed of two falling objects that differed in size or weight, their prior beliefs that the objects would fall at different speeds affected both their observations and conclusions from the data. Thus, after viewing two objects released at the same time from the same height, many children claimed that they had observed one object fall faster than the other, despite the fact that they fell at the same rate (Chinn & Malhotra, 2002).

Many investigations of children's data interpretation have done so within knowledge-rich domains (e.g., Echevarria, 2003; Metz, 2004). However, children often encounter situations about which they have very limited background knowledge. In these

---

This work was supported in part by grants from the National Institute of Child Health and Human Development (HD25211) to David Klahr, and from the National Institute of Mental Health training Grant T32 MH19102. Thank you to the parents, teachers, and principals at St. Teresa of Avila, Sterrett Classical Academy, Greenfield Elementary, Sacred Heart Elementary, Schiller Classical Academy, Colfax Elementary, and the Frick Academy for their invaluable assistance. We also owe many thanks to Anne Siegel, Audrey Russo, and Jen Schnakenberg for assistance with data collection and coding; to Steve Tibbets and Julie Heath for assistance with SAS programming; and to David Klahr, Corinne Zimmerman, and three anonymous reviewers for helpful comments on earlier drafts. Portions of the preliminary results from Experiment 1 were summarized in Masnick, A. M., Klahr, D., & Morris, B. J. (2007). Separating signal from noise: Children's understanding of error and variability in experimental outcomes. In M. Lovett & P. Shah (Eds.) *Thinking with data* (pp. 3–26). New York: Lawrence Erlbaum. In addition, portions of these results were reported at the meetings of the American Educational Research Association (2002 and 2004), the Society for Research in Child Development (2003 and 2005), and the Cognitive Science Society (2002 and 2004).

Correspondence concerning this article should be addressed to Amy M. Masnick, Department of Psychology, Hofstra University, Hempstead, NY 11549. Electronic mail may be sent to amy.m.masnick@hofstra.edu.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7904-0015

situations, data provide a rich source of information from which to draw conclusions about a phenomenon. What is not known is whether children attend to and use particular diagnostic features of data. That is, given a set of observations, are there properties of data sets—unrelated to knowledge of the domain—that help reasoners induce trends from data sets without using formal statistics?

We suggest that two diagnostic features of data patterns may play an important role in reasoning: sample size and variability. Sample size refers to the number of observations within a data set. Considering the golfing example mentioned previously, someone using sample size as a factor to assess differences between two golfers would likely be more confident about drawing conclusions with more data available (e.g., three hits per golfer vs. one hit per golfer). Variability refers to the degree of similarity or dissimilarity among values within and between the set of observations. Returning to the golfing example, *within-group* variability is a measure of differences in the outcomes of drives made by the same golfer, whereas *between-group* variability is a measure of differences in drives between golfers. A reasoner using variability characteristics of data would be more confident of conclusions drawn from data as the variability within a set of observations decreases and variability between a set of observations increases.

Reasoning About Data in Category Induction and Scientific Reasoning Tasks

In category membership tasks, there is evidence that elementary school children recognize sample size and variability. Gutheil and Gelman (1997) found that although adults used variability and sample size in combination or independently, 8- and 9-year-old children did not differentiate between sets of exemplars that differed only in sample size or in variability but did differentiate when both sample size and variability differed between sets. In related work, Jacobs and Narloch (2001) asked children in Grades 1, 3, and 5 to predict how common a trait would be in a given population based on a sample that varied in number (sample size) and homogeneity (within-group variability). Prior knowledge about variability affected children's predictions more than sample size. Strong prior assumptions about the phenomena in question influence reasoning about data characteristics.

There is also evidence that children attend to sample size and data variability in scientific reasoning contexts. Piaget and Inhelder (1951/1975) investigated children's assessments of the probability of a spinner landing in a series of possible locations. By middle childhood, children's judgments about future landing sites were related to the number of previous observations, indicating some understanding of sample size (i.e., the law of large numbers, or the idea that larger sample sizes make one more confident about generalizing to a population). Klaczynski and Aneja (2002) also demonstrated that children as young as 7 years old applied the law of large numbers in inferring new gender stereotypes by being more confident when shown a larger sample than a small one.

A very common scientific reasoning task involves drawing conclusion based on the evaluation of covariation data (e.g., Koerber, Sodian, Thoermer, & Nett, 2005; Kuhn, Amsel, & O'Loughlin, 1988; Shaklee & Paszek, 1985). Interpreting covariation data is also influenced by data characteristics. Koslowski, Okagaki, Lorenz, and Umbach (1989) asked sixth-grade, ninth-grade, and college participants whether a target factor was likely to cause a particular effect (e.g., whether using a gasoline additive leads to worse gas mileage). Information presented to participants varied in sample size and in whether there was evidence of covariation. Without covariation information, all participants differentiated between large and small sample sizes; however, with covariation information, only the college students used sample size information. Thus, there is evidence that sixth and ninth graders have some difficulty integrating multiple characteristics of the data. At the same time, even sixth graders can use sample size in the absence of other information to infer cause. Koslowski (1996, Experiment 10) also reported that when more instances of covariation were present, participants were more likely to rate the association between two variables as causal than when there was only one instance of covariation.

Children also attend to variability within and between data sets. When second and fourth graders rolled two balls down two ramps and measured the distance each traveled, they distinguished between the small differences expected in individual data points (within-group variability) and larger differences expected between groups (between-group variability; Masnick & Klahr, 2003). The expectation that there will be some variation in precise data points indicates an understanding that many factors affect the outcome of an experiment, even when these small differences may not affect overall conclusions.

Yet, there is evidence that children have difficulty understanding what this variability indicates. Lubben and Millar (1996) showed 11-, 13-, and 15-year-olds two sets of results with equal means but different amounts of variation. Participants were asked whether one set of results was more trustworthy or if they were equally trustworthy. About half (47% – 55%) of all participants said that one data set was more trustworthy than the other. However, only about 20% of 11-year-olds justified their choice by referring to variation in the data. Forty-four percent of 13-year-olds and 48% of 15-year-olds justified their choice by referring to variation in data, indicating an understanding that data variability is a component of assessing reliability (Lubben & Millar, 1996). Thus, even into high school, a sizable number of students have difficulty using variability information in drawing conclusions.

Studies of covariation detection, in effect, ask participants to assess an "intuitive chi-square" estimate: How likely is it that the observed pattern of frequencies occurred by chance? Although much of the research on the evaluation of evidence has taken the form of covariation matrices concerning the frequency of the presence and absence of putative causes and effects (i.e., an intuitive chi-square, as in Shaklee & Paszek, 1985), in both real-world contexts and the science classroom, children must evaluate the differences in a range of *quantitative* data. Zimmerman (2000) noted that to determine if a specific antecedent is linked with an outcome, one compares the number of times the outcome occurs or does not occur when the antecedent is present. If the former is much larger than the latter, one can generally conclude that there is a relationship between the two. However, she observed that "it is not clear *how large* the difference must be in order to conclude that the two events are related" (p. 115). The question is then, How do students reason when comparing sets of quantitative data?

In addition to the evidence that elementary and secondary school children attend to features of data in some contexts, there is also evidence that very young children have some fundamental understanding of relative quantity. Even 6-month-old infants can accurately compare two quantities (see Feigenson, Dehaene, & Spelke, 2004, for a discussion), and 5-year-old children are equally adept at determining which quantity is larger when presented Arabic numerals or arrays of dots (Temple & Posner, 1998). Thus, it seems possible that some of the skills required for effective data comparison are implicit and not the result of direct training. It is possible that children use a relatively automatic process of quantity comparison. We follow Rubinsten, Henik, Berger, and Shahar-Shalev (2002) in suggesting that "automatic" refers to processing that requires no conscious monitoring after commencing. The process that operates may be an "intuitive" statistical test on data patterns. We use the term "intuitive" to convey the automaticity of the process.

Clearly, comparing sets of data requires processing beyond that of simply detecting differences in magnitude. A second set of processes is needed to process exact differences (e.g., perform calculations), to examine characteristics of data, and to integrate this information with domain knowledge. Inhelder and Piaget (1958) suggested that a pattern of fixating on one feature instead of looking at features in concert was the source of young children's difficulty in solving conservation tasks. If children do focus initially on a small number of features, then we would expect to see age-related increases in the number of characteristics to which they attend. Additionally, children would likely develop the ability to compare sets along multiple dimensions before they could explicitly access knowledge to consider each variable independently within the set. It is likely that these improvements are due to many factors including acquisition of new strategies, increases in domain knowledge, and increases in processing capacity (Halford, Cowan, & Andrews, 2007).

Present Study

To our knowledge, there has been no systematic investigation of the characteristics of data to which children attend and to what extent these characteristics influence their judgments about phenomena. The present study was designed to examine which characteristics of data guide inferences when comparing data sets. First, we predicted that sample size is likely to play an important role in reasoning. Second, we predicted that in drawing conclusions about comparative data, two characteristics that indicate the *amount of variation* in the data are key: within-group variability (i.e., the variability of the data points relative to mean) and between-group variability (i.e., variability of data points in each of two data sets relative to one another). This information about variability can be assessed increasingly well with a larger sample size of data points. Finally, we predicted that with age, children will become better at using multiple characteristics of data in concert and will have more explicit understanding of the importance of these characteristics for making judgments about quantitative data.

We chose to work with third- and sixth-grade students to explore developmental changes through elementary school because past work on reasoning with data has often examined children in these age groups (e.g., Gutheil & Gelman, 1997; Inhelder & Piaget, 1958; Jacobs & Narloch, 2001; Koslowski et al., 1989), as they become exposed to experimental data in school. Our pilot data indicated that third graders were the youngest age group in which all participants could consistently perform our data interpretation tasks. A 3-year age difference allows for a large enough difference to see change if it is apparent. We chose a college student comparison population because it is clear that these skills continue to develop with experience and age.

In two experiments, children and college students were presented with sets of paired data and asked to draw conclusions about differences between the sets. In Experiment 1, data were presented with one of two framing stories that limited the amount of background knowledge that could be brought to bear in drawing conclusions about the data. The data presented varied systematically in number of data points presented, within-group variability (operationalized as a dichotomous variable, with two sizes of standard deviations relative to the mean), and between-group variability (operationalized as the number of pairs of data in which the data point from the column with the lower mean was higher than the data point from the column with the higher mean).

When studying reasoning about any topic, it is important to consider both implicit and explicit evidence of knowledge used. Nisbett and Wilson (1977) argued that explicit justifications for behaviors are often distinct from actual reasons for behavior and that the salience of a stimulus is the key determinant of it being used as a justification. Thus, a full understanding of the cognitive process involves looking at both patterns of conclusions drawn from data and explicit rationales for these patterns. Participant responses were analyzed to examine how each of the three characteristics of data (sample size, between-group variability, and within-group variability) was used in implicit and explicit reasoning.

Experiment 1

Method

Participants. Thirty-nine third graders (mean age = 9.1 years, range = 8.2 – 10.3 years), 44 sixth graders (mean age = 11.9 years, range = 11.2 – 12.8 years), and 50 college undergraduates (mean age = 20.2 years, range = 18.1 – 23.7 years) participated in this study. College students were recruited from undergraduate psychology courses, and younger participants were recruited from letters sent to parents at four elementary and middle schools in the northeast United States. The third-grade sample consisted of 92% White students, 5% Black students, and 3% Hispanic students. The sixth-grade sample consisted of 82% White students, 2% Asian students, and 16% Black students. The college student sample consisted of 54% White students, 40% Asian students, 2% Black students, and 4% Hispanic students.

Procedure. All participants were interviewed individually. Participants were randomly assigned to one of two cover story conditions: one in which robots were the source of the data and one in which athletes (people) were the source of the data. In the robot condition, each participant was read the following information:

> Some engineers are testing new sports equipment. Right now, they are looking at the quality of different sports balls, like tennis balls, golf balls and baseballs. For example, when they want to find out about golf balls, they use a special robot launcher to test two balls from the same factory. They use a robot launcher because they can program the robot to launch the ball with the same amount of force each time. Sometimes they test the balls more than once. After they run the tests, they look at the results to see what they can learn.

In the athlete condition, we used an isomorphic cover story in which two athletes were trying out for one slot on a sports team. The coaches asked the participants to perform certain tasks (e.g., hit a golf ball as far as possible) to assess which athlete would be better for the team. For example, participants in the athlete condition saw the following story: "Harriet and Joan are trying out for the soccer team. The soccer coach asked them to kick a soccer ball four times. The coach measured how far the soccer ball went each time it was kicked." This cover story was designed to see if adding information about a potential source of variability (human error) would change participants' responses. Thus, the robot condition presented a cover story that minimized potential sources of variability, whereas the athlete condition provided potential sources of variability.

After reading the cover story, the participants were shown a series of 14 data sets, one at a time. For each data set, there were data either for two balls of the same type, with no distinguishing characteristics (e.g., Baseball A and Baseball B), or for two athletes, with no information other than first names (e.g., Alan and Bill). The 14 balls tested were golf balls, racquetballs, basketballs, soccer balls, ping pong balls, footballs, volleyballs, baseballs, tennis balls, marbles, hockey pucks, kickballs, softballs, and bowling balls. Participants were randomly assigned to one of two orders of presentation. There were no order effects; therefore, results are collapsed across order. In the athlete condition, different names were used for each data set to prevent any carryover knowledge effect.

For each data set, there were one, two, four, or six pairs of data. Each page contained two columns of data: one column listing the distance the first ball traveled and one listing the distance the second ball traveled. (See Table 1 for specific examples of the different data characteristics.)

The 14 data sets varied in (a) sample size, (b) between-group variability (i.e., reversals in which column's number was higher), and (c) within-group variability (i.e., high or low variability relative to the means). Each participant evaluated 14 comparisons, with eight trials including no reversed pairs (sample sizes 1, 2, 4, and 6) and six trials including one or two reversed pairs (sample size 4 with one reversed pair, and sample size 6 with one and two reversed pairs). Half of the trials had high within-group variability, in which the standard deviation of the data set was 15% – 20% of its mean, and half had low within-group variability, in which the standard deviation of the data set was less than 2% of its mean. Each of the 14 trials tested a different type of sports ball.

For each data set, participants were asked first what the engineer or coach could find out as a result of this information and to explain any reasons for their answer. Next, they were asked how sure they were about these conclusions. To answer the questions about sureness, participants were offered a 4-point scale from which to select their answer, choosing among *not so sure*, *kind of sure*, *pretty sure*, and *totally sure*.

Measures. For each question in which participants were asked to report their conclusions and predictions, and their confidence in their answers, participants first answered yes or no and then rated their confidence on a 4-point Likert scale. These two responses—yes – no and sureness level—were combined into a single 7-point ordinal variable: *totally sure there is a difference*, *pretty sure there is a difference*, *kind of sure there is a difference*, *not so sure* (regardless of yes or no answer), *kind of sure there is no difference*, *pretty sure there is no difference*, *totally sure there is no difference.* Past research has demonstrated that the children as young as 8 years old have been able to understand and work with such a scale (Masnick & Klahr, 2003). Participants were asked to provide reasons for their initial conclusions and final predictions of relative position. These reasons were coded for mention of data characteristics (such as sample size or variability) or mechanism (such as a property of the ball that could have affected the results).

Table 1
Examples of Data Sets in Experiment 1

Example 1: Six data pairs with one reversed data pair (five of six times Basketball B goes farther), high within-group variability, robot condition

Basketball A	Basketball B
56 ft	74 ft
65 ft	83 ft
75 ft	57 ft
49 ft	66 ft
48 ft	67 ft
64 ft	82 ft

Example 2: Two data pairs with no reversed pairs, low within-group variability, athlete condition

Joan	Harriet
372 ft	383 ft
363 ft	374 ft

Results

The main research question was whether different types of data patterns and different contexts affect participants' decisions about, and confidence in, their judgments of whether there is a difference between the two groups of data presented. Decisions about whether the data sets differed were operationalized by participants' responses to whether they thought the engineer or coach could be sure that there was a difference between the two groups. In addition, participants' reasons for the responses were analyzed.

Use of data characteristics to judge whether there was a difference between the data sets. For each assessment, sample size, between-group variability (number of reversed pairs), and within-group variability (spread) were within-subjects variables, and cover story was a between-subjects variable. There were no gender differences. In addition, for the college students, there were no differences in ratings for those who had taken different numbers of statistics classes. Thus, these two variables were not considered in later analyses. Because the data were not set up in a full factorial design, each data characteristic was analyzed separately. Of those who said that one ball traveled farther than the other, third graders were accurate in stating which ball went farther 92.4% of the time, sixth graders were accurate 98.8% of the time, and college students were accurate 99.5% of the time.

A 4 (sample size: 1, 2, 4, or 6) × 3 (age: third grade, sixth grade, or college) × 2 (cover story: robot vs. athlete) mixed analysis of variance (ANOVA) was used to assess whether sample size influenced confidence judgments in cases with no between-group variability. There was a main effect of sample size, $F(3, 120) = 20.68$, $p < .001$, partial $\eta^2 = .34$; a main effect of age, $F(2, 122) = 44.52$, $p < .001$, partial $\eta^2 = .42$; no main effect of cover story, $F(1, 122) = 0.70$, $p > .1$; and a Sample Size × Age interaction, $F(6, 240) = 17.63$, $p < .001$, partial $\eta^2 = .31$ (Figure 1). Tukey's honestly significant difference (HSD) post hoc tests indicated that college students' ratings were significantly lower on average than the third and sixth graders'. College students' sureness ratings increased dramatically with sample size, $F(3, 44) = 50.59$, $p < .001$, partial $\eta^2 = .78$, and sixth graders' ratings increased a small but significant amount with increased sample size, $F(3, 38) = 3.33$, $p = .030$, partial $\eta^2 = .21$. Third graders, in contrast, showed a slight trend of *decreased* sureness ratings with increased sample size, $F(3, 34) = 2.26$, $p = .099$, partial $\eta^2 = .17$. There were no interactions with the cover story. In sum, at all age levels, it was apparent that participants were sensitive to differences in sample size. For sixth graders and college students, this observation tended to increase ratings of sureness; for third graders, it tended to decrease ratings.

In addition to sensitivity to sample size, we also explored sensitivity to the presence of reversed data pairs (between-group variability) by examining only the data sets with six pairs of data. This pattern occurred when there were no reversed data pairs, one reversed pair, and two reversed pairs. A 3 (between-group variability: zero, one, or two pair reversals) × 3 (age: third grade, sixth grade, or college) × 2 (cover story: robot vs. athlete) mixed ANOVA demonstrated no effect of cover story, $F(1, 123) = 0.56$, $p > .1$. There was a main effect for between-group variability (lower sureness ratings with more variability), $F(2, 122) = 53.15$, $p < .001$, partial $\eta^2 = .47$; a main effect for age (higher average ratings for third graders than for sixth graders and college students), $F(2, 123) = 9.59$, $p < .001$, partial $\eta^2 = .14$; and a Variability × Age interaction, $F(4, 244) = 4.70$, $p = .001$, partial $\eta^2 = .07$ (Figure 2). At all age levels, participants' mean sureness ratings decreased with more reversed data pairs between the two groups of data, with the effect stronger with increasing age: for third graders, $F(2, 34) = 3.51$, $p = .041$, partial $\eta^2 = .17$; for sixth graders, $F(2, 40) = 14.42$, $p < .001$, partial $\eta^2 = .42$; and for college students, $F(2, 46) = 68.15$, $p < .001$, partial $\eta^2 = .75$.

We also assessed differences in within-group variability (data spread) on ratings of confidence that there was a difference between the two data sets. Participants saw two data sets with each sample size and between-group variability combination used (e.g., two data pairs with no reversed pairs; six data pairs with one reversed pair). One of these data sets had high within-group variability (standard deviation between 15% and 20% of mean) and one had low within-group variability (standard deviation less than 2% of mean). We calculated a paired t test for each sample size/between-group variability combination participants saw (directly comparing sureness ratings from data with high and low within-group variability). Third and sixth graders showed no evidence of differentially rating data sets based on within-group variability. College students had an inconsistent pattern, in which three of the six comparisons reached significance (for four pairs with no reversed pairs, for six pairs with one reversed pair, and for six pairs with two reversed pairs). Surprisingly, in each of these comparisons, participants were more sure in the case with *higher* variability. Possible explanations for this pattern are proposed in the Experiment 1 Discussion section.

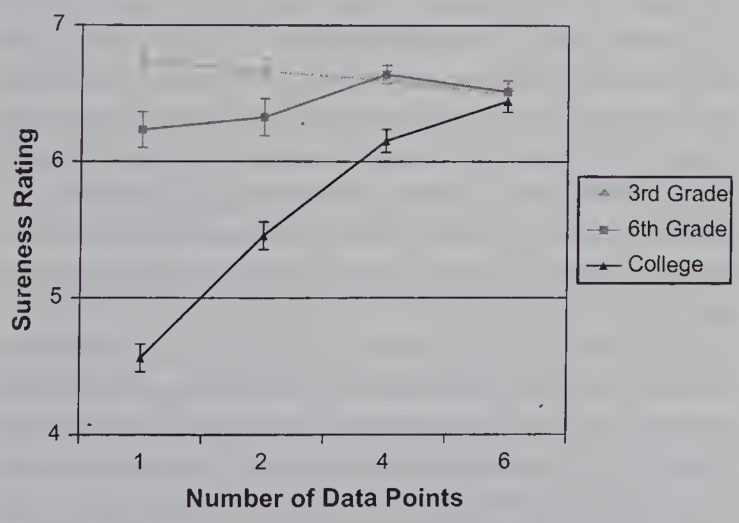


Figure 1. Experiment 1: Participants' sureness ratings at each grade with one, two, four, and six pairs of data but no reversed data pairs (collapsed across cover story).
Note. Error bars represent standard error of measurement.

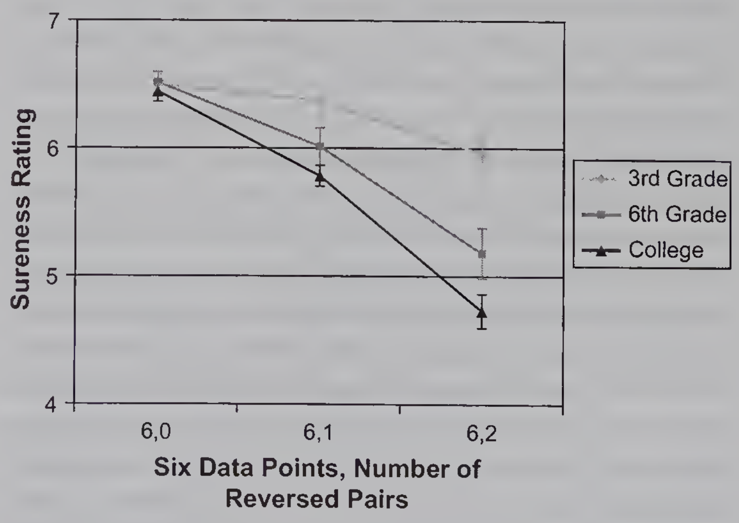


Figure 2. Experiment 1: Participants' sureness ratings at each grade with six pairs of data, by level of between-group variability in the data (collapsed across cover story).
Note. Error bars represent standard error of measurement.

When people evaluate differences between sets of data, they do not focus on only one characteristic. For example, in computing formal statistics, a *t* statistic takes account of both sample size and variability when testing for significant differences between means. The next analysis explored the possibility that our participants used an "intuitive *t* test" in which a combination of the number of data points and the variability of the data points were both considered in evaluating differences between means. Although these are small data sets, we calculated the independent *t*-test statistic for each set of data presented to the participants to examine the trends (with the obvious exception of the two sets with one pair). These 12 *t* values ranged from 0.88 to 4.18. Next, we correlated the *t* statistics for the presented data with the mean sureness rating participants gave assessing whether there was a difference between the two sets of data, by grade. An interesting pattern emerged. Third graders' ratings were not significantly associated with the independent *t* statistic ($r = .388, p = .213$), sixth graders' ratings showed evidence of a strong relationship ($r = .632, p = .028$), and college students showed evidence of a very strong relationship ($r = .873, p < .001$).

Reasons offered. Participants offered explicit justifications for confidence or lack thereof in their conclusions and for the predictions they made about which ball would go farther on a subsequent trial. Reasons given were coded for mention of the manipulated data characteristics: (a) sample size, (b) between-group variability (whether the data sets included pair reversals or not), and (c) within-group variability (spread of the data). We also noted mention of (d) a trend in the data and (e) the magnitude of the difference between the two data sets. Additionally, explanations of how the outcome could have occurred were coded, including (f) a property of the ball that affected the results, (g) a property of the robot or athlete that affected the results, or (h) a property of the environment—other than the ball, robot, or athlete—that affected the results. The categories were not mutually exclusive, and participants had 28 opportunities to explain their reasoning (twice for each of the 14 data sets). Participants' responses were coded by two independent coders, whose agreement level was 85%. All discrepancies were resolved via discussion.

Table 2 presents a summary of these results. An overwhelming majority of the responses were in reference to the data and not to theoretical issues such as properties of the ball or robot/athlete. Many participants mentioned the specific data characteristics that were manipulated, including sample size (e.g., "They only tested it two times so they can't really be sure"), between-group variability characteristics such as reversed pairs ("All but once A went farther than B") or no reversed pairs ("A went farther than B every time"), and within-group variability characteristics such as variability within the column ("All the numbers in A are close together").

Table 2

Experiment 1: Percentage of Participants at Each Age Who Mentioned Each Justification for Their Sureness

	Third grade	Sixth grade	College
Data responses			
Sample size**	10	39	100
Reversed pairs*	56	70	90
No reversed pairs**	8	14	62
Spread of data**	3	15	44
Overall trend in data	92	93	100
Single pair "trend"**	23	0	0
Magnitude of differences**	46	89	92
Magnitude of single pairs	5	2	0
Mechanism responses			
Ball property	15	20	18
Robot/athlete property	41	37	30
Environment	3	2	2

Note. Age differences: $*p < .05$. $**p < .01$.

In addition, other characteristics not manipulated were also mentioned. The most frequently mentioned data characteristic was a trend in the data (e.g., "Five out of six times A went farther"; "B went farther most of the time"). Some third graders (23%) explicitly referred to only one pair of the data set, effectively ignoring the remaining data in their justifications; we called this type of response a "single-pair trend." No sixth graders or college students justified their reasoning this way. Another data characteristic mentioned was the magnitude of differences between the two columns of data (e.g., "A went much farther than B"). In general, college students used a much wider range of responses than younger children, with nearly all of them mentioning sample size at least once. However, all participants, with the exception of 2 third graders, mentioned at least one data characteristic at some point during the study.

In addition to discussions of data, we also explored whether participants offered causal or mechanistic explanations as justifications for their conclusions (Table 2). We coded explanations for a mention of characteristics of the robot (e.g., "Maybe the robot was breaking down after it threw Ball A"), characteristics of the athlete (e.g., "Maybe Person A was having a good day"), characteristics of the ball (e.g., "Maybe Ball B was more aerodynamic"), and characteristics of the environment (e.g., "Maybe there was wind blowing when Ball A was being thrown"). Across all trials, about half of the participants mentioned at least one mechanistic explanation, with no age differences (51% of third graders, 56% of sixth graders, and 42% of college students mentioned such an explanation).

One area in which a cover story difference might be expected is in use of these mechanistic justifications. Indeed, use of these mechanistic justifications did vary considerably by condition: Sixty-two percent of participants in the athlete condition said that a property of the athlete was a reason for at least one outcome, whereas 10% of those in the robot condition suggested a property of the robot as a reason for an outcome at least one time, $\chi^2(1, N = 133) = 38.19, p < .001$. In contrast, 6% of participants in the athlete condition and 29% of participants in the robot condition suggested a property of the ball as a potential explanation for the outcome at least one time, $\chi^2(1, N = 133) = 12.16, p < .001$. Mentions of other environmental mechanisms such as wind only occurred three times, twice in the robot condition and once in the athlete condition.

Discussion

When sample size, between-group variability, and within-group variability were manipulated, we found that sample size and the between-group variability influenced participants' reasoning but that within-group variability had no clear effect. In addition, older participants were able to make more accurate distinctions between data sets, indicating better knowledge of data characteristics individually and in concert. Children were also able to talk about data characteristics as early as in the third grade, demonstrating an early recognition of the importance of these concepts, but there were significant age-related increases in the number of data characteristics described in their explanations.

Although prior research indicates sensitivity to the law of large numbers at as early as age 7 (e.g., Klaczynski & Aneja, 2002), only college students indicated a clear appreciation of the law of large numbers by increased confidence ratings with increased sample size. Sixth graders showed a mixed pattern, and a sizeable number of third graders were less sure about conclusions when there were more data points—though this result does suggest sensitivity to the number of data points presented. These students may still be looking for a single correct answer and find multiple data points to be more confusing than informative.

In addition, there is evidence that in all three age groups, children pay attention to between-group variation in the data, one of the more complicated statistical concepts. Between-group variability in data is clearly salient to children even when the data are not couched in a strong theoretical context. Though this concept is relatively difficult in formal statistics, it may be implicitly provided by the presentation format of the data itself. For example, when scanning numbers presented in lists, when all numbers are visible simultaneously, and recall is not being tested, children may encode information about variability while encoding the values themselves. Thus, with age-related increases in processing capacity and speed as well as more optimal strategy use (e.g., Gathercole, Pickering, Ambridge, & Wearing, 2004; Kail, 2007), participants showed improvement in their ability to use information about sample size and between-group variability and to use the characteristics together in the form of an intuitive *t* test.

Participants did not seem to respond to changes in within-group variability, our manipulation of the standard deviation relative to the mean, though college students showed a small trend toward being more sure in cases with greater variability. It is possible that this characteristic was simply not salient or believed to be important and thus did not influence most responses. It is also possible that the lack of effect for children, and small reverse effect for college students, occurred because variability was confounded with data magnitude, such that the data with low variability relative to the mean consisted entirely of three-digit numbers, whereas the data with high variability relative to the mean consisted entirely of two-digit numbers. This confounding may have led to college participants simply rating themselves as more confident when reasoning about two-digit numbers instead of the more challenging three-digit numbers. In addition, with small sample sizes, the spread of the data may have been less noticeable.

One of our more intriguing findings was that participants appear to become increasingly adept with age at performing intuitive *t* tests on the data sets, that is, in using the data characteristics in concert. The lack of association of third graders' ratings with the *t*-test values of the data sets may be due to their consistently high confidence ratings. However, for the older participants, it indicates that even when participants were not systematically using the individual characteristics, they were able to use them together to draw conclusions, regardless of their level of explicit access to this information. The fact that the college students showed a small trend in being more confident with increased within-group variability while demonstrating strong evidence of drawing conclusions closely aligned with *t*-test results may indicate that for these data sets, the overall within-group variability was relatively small (even with our high-variability manipulation). Therefore, reasoning that focused on sample size and between-group variability may have been a reasonable way to draw conclusions about these data.

The participants not only used data characteristics in drawing conclusions but also justified their conclusions explicitly. In doing so, nearly every participant mentioned at least one data characteristic, though the college students referred to a wider range of characteristics and referred to these characteristics more frequently. The third graders appeared to be much less explicitly aware of concepts such as sample size and within-group variability, and the majority of their data-related comments were about trends in the data. Nonetheless, about half of the third graders referred to between-group variability and about half referred to the magnitude of differences between groups. The mention of these characteristics at all at such a young age indicates an early awareness that these features are important ones to consider in drawing conclusions from data.

In addition to the data features, even when we set up a situation with no clear mechanistic explanation, about half of the participants at each age brought their background knowledge to bear and mentioned potential mechanisms for the results that were not explicitly mentioned in the cover stories. Therefore, all students did not evaluate the data in a theory-neutral way. For some, predictions appealed to patterns in the previous data sets. For other students, predictions appealed to theoretical reasons for *why* such patterns were found.

One possible limitation of this study is that the presentation of data might have influenced reasoning by either providing too much information at once or guiding interpretation. Format of problem presentation has been shown to affect children's ability to solve math problems (e.g., Klein & Bisanz, 2000). If this amount of information exceeds processing limits, participants might use nonoptimal strategies for analysis. For example, instead of comparing the entire range of data at once, participants might draw a conclusion from the last set of numbers alone, ignoring the remaining data. In addition, for each trial in Experiment 1, participants were presented two data sets and asked to compare them. By presenting the data in this way, it is possible that participants would only compare *pairs* of data points instead of comparing the entire columns of data. To examine these issues, we conducted a second experiment.

Experiment 2

In the follow-up experiment, we explored how style of data presentation might influence children's and adults' conclusions. We made several specific changes to the methodology from Experiment 1. We changed the data presentation style so that the data were presented incrementally instead of simultaneously. This change in presentation allowed us to reduce the information processing burden and to compare the effect of presentation style on participants' processing of the data. Specifically, we wanted to explore whether the change from simultaneous to paired presentation would highlight different data characteristics (e.g., sample size) and obscure others (e.g., between-group variability). In addition, would a presentation format that prevents direct paired comparison lead to differences in reasoning about data characteristics? If understanding of data characteristics is robust, then the presentation format should have a minimal effect on using the characteristics to draw conclusions. However, if participants were only looking at subsets of the data, then we would expect the pairwise condition to lead to more confidence in evaluations derived from smaller samples. If participants were relying on individual pairwise comparisons as part of their reasoning, then the pairwise condition would likely lead to reasoning more clearly based on between-group variability than the column condition, in which the opportunities for such comparisons were limited. We also predicted that the two formats would lead to increased attention to the data characteristics, particularly to sample size, as the feature that most obviously changed between iterations.

We made other minor modifications in response to observations of participants in Experiment 1. We wanted to find out how likely participants were to want the engineers to test the balls again, an assessment of whether participants believed that they needed additional data to draw a conclusion. From the Experiment 1 cover story, one could argue that there is always a reason to be conservative in conclusions and state that there is not really evidence for a difference (a bias toward a Type II error rather than Type I error). Therefore, we modified the cover story to indicate more clearly that there are costs to being wrong in either assuming that balls go the same distance when they do not or assuming that they do not when in fact they do. This change was to reduce the possibility of participants suggesting simply producing and testing more balls without first considering the available data.

In addition, we changed some of the specific questions to be more explicit about exactly what we were looking for. In Experiment 1, we asked participants what the engineer had learned from testing, and nearly all participants gave answers about which ball went farther. If they did not, the experimenter probed for that answer directly. In Experiment 2, we changed the question to ask explicitly which ball went farther.

Finally, we added a practice problem that included three presentations of data and one reversed data pair to familiarize all participants with the nature of the task. The practice problem served to introduce participants to the notion that they would be seeing more data after the first trial and also that pair reversals in the data sometimes occur.

Method

Participants. Twenty-two third-grade students (mean age = 9.0 years, range = 8.2 – 10.6 years), 29 sixth-grade students (mean age = 11.9 years, range = 11.1 – 13.8 years), and 50 undergraduate students (mean age = 19.8 years, range = 17.8 – 23.6 years) participated in this study. None of these students participated in Experiment 1. The children were students in two elementary schools and two middle schools in the Northeast United States. The third-grade sample consisted of 54% White students, 21% Asian students, 18% Black students, and 9% Hispanic students. The sixth-grade sample consisted of 56% White students and 44% Black students. The third- and sixth-grade students were from different regions in the same metropolitan area than those who participated in Experiment 1. The college student sample consisted of 62% White students and 38% Asian students.

Procedure. Because there were no differences in the main conclusions drawn based on cover story, all participants in Experiment 2 were shown the robot cover story used in Experiment 1, with following paragraph added:

> It's very important that the engineers can find out whether the balls are the same or different. Only if the balls are the same and go the same distance can the factory sell them to make money. If they aren't the same then the engineers need to remake the balls, which is expensive and time-consuming. So the engineers need your help in looking at the results and deciding if they can be sure whether the balls are the same or different.

Participants were presented with three data sets, each with a different ball type, in one of two presentation formats. The three data sets were identical to those used in Experiment 1, and each included six pairs of data. One set had no reversed pair, one had one reversed pair (the third data pair was greater in Column B than in Column A), and one had two reversed pairs (the second and fifth data pairs were greater in Column A than in Column B). The order of presentation was counterbalanced using a Latin Square design, such that one third of the participants in each age group saw each data set first. All the data used in this study were from the high-variability (two-digit) condition from Experiment 1, as we found no clear effect of within-group variability on reasoning.

In the pairwise condition, participants saw data presented in two columns. First, they were shown one pair of data points, then two pairs, then four pairs, and then six pairs. After each presentation of data, participants were asked if there was a difference between the two balls, how sure they were of whether there was a difference and why, and whether they thought that the engineers should test the balls again. Examples of this condition are shown in Table 3.

Table 3
Example of Data Sets in Experiment 2

Examples of presentation formats shown to participants in the pairwise condition

Basketball A	Basketball B
74 ft	56 ft

Basketball A	Basketball B
74 ft	56 ft
83 ft	65 ft

Examples of presentation formats shown to participants in the column condition

Tennis Ball A	Tennis Ball B
49 ft	64 ft
40 ft	
39 ft	
51 ft	
50 ft	
38 ft	

Tennis Ball A	Tennis Ball B
49 ft	64 ft
40 ft	52 ft
39 ft	
51 ft	
50 ft	
38 ft	

In the column condition, participants saw six data points in one column and one in the other column. They were asked the same questions as in the pairwise condition and were then presented with additional data points in the second data column (one, two, four, and six data points at a time). This condition was included to see if reasoning changes when pairwise comparisons are less salient. Examples of this format are shown in Table 3. Questions asked in both formats were identical, with no explicit mention of the format in the questions.

Results

Judgments of whether there was a difference between the data sets. As in Experiment 1, responses—yes–no and sureness level—were combined into a single 7-point ordinal variable. Because there was no direct assessment of whether participants believed that the ball in Column A or Column B went farther, there was no accuracy measure. Participants stated only whether they thought that there was a difference between the two variables.

To examine the effect of sample size alone, we examined the data set in which reversed pairs were not included. Thus, we compared judgments of whether there was a difference between data sets when participants saw one, two, four, or six pairs of data or one column with six data points and a second with one, two, four, or six data points. These data points were all presented in the same scenario, and data points in one column were always larger than those in the other column. We found that sample size influenced participants' assessments of differences between the data sets. Age and condition were between-subjects variables, and sample size was a within-subjects variable. A 3 (age: third grade, sixth grade, or college) × 2 (condition: pairwise vs. column) × 4 (sample size: 1, 2, 4, or 6) mixed-model ANOVA was performed. We found no significant effect of age, $F(2, 95) = 1.73, p > .1$; no effect of condition, $F(1, 95) = 1.14, p > .1$; and a significant effect of sample size, $F(3, 93) = 4.52, p = .005$, partial $\eta^2 = .13$. There was also an Age × Sample Size interaction, $F(6, 186) = 3.87, p = .001$, partial $\eta^2 = .011$. Simple effects analysis to understand the nature of the interaction indicate that college students demonstrate a large, significant increase in sureness with increasing sample size in the pairwise condition, $F(3, 21) = 7.40, p = .001$, partial $\eta^2 = .51$, and in the column condition, $F(3, 23) = 17.37, p < .001$, partial $\eta^2 = .53$, but that third and sixth graders do not ($Fs < 1.0$). In addition, there was a significant Sample Size × Condition interaction, such that the paired condition leads to a linear trend of increasing sureness ratings with increased sample size, whereas the column condition does not, $F(3, 93) = 4.97, p = .003$, partial $\eta^2 = .14$. See Figure 3 for means for each age and sample size, by condition.

Between-group variability also influenced assessments of differences between data sets. The ratings of whether there was a difference between groups on the sixth trial (when all data had been presented) were compared across stories in a mixed-model ANOVA. Each participant saw one story with no reversed pair, one with one reversed data pair, and one with two reversed data pairs. A 3 (age: third grade, sixth grade, or college) × 2 (condition: pairwise vs. column) × 3 (between-group variability: zero, one, or two reversed data pairs) ANOVA was performed. We found no effect of age, $F(2, 95) = 0.74, p > .1$; an effect of condition, $F(2, 95) = 5.59, p = .020$, partial $\eta^2 = .06$; and an effect of between-group variability, $F(2, 94) = 12.60, p < .001$, partial $\eta^2 = .21$. On average, there were higher ratings of a difference in the pairwise condition and higher ratings with less between-group variability (i.e., with fewer reversed pairs). There was also a significant Age × Between-Group Variability interaction, $F(4, 188) = 3.87, p = .005$, partial $\eta^2 = .08$. Examining the Age × Variability interaction, only the college students reduced their sureness ratings considerably when there was more between-group variation, in both the pairwise condition, $F(2, 22) = 18.94$,

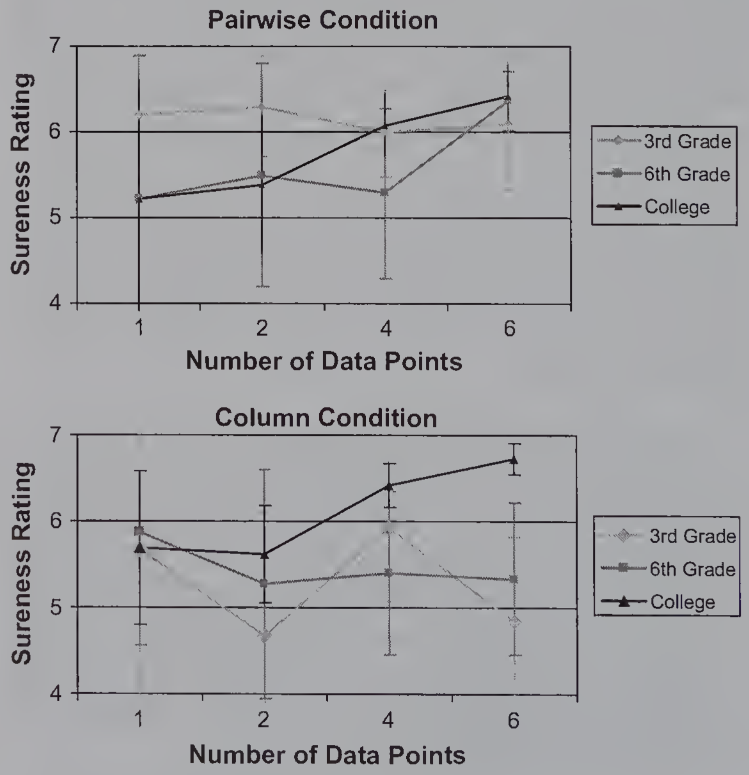


Figure 3. Experiment 2: Participants' sureness ratings at each grade with one, two, four, and six pairs of data but no reversed data pairs (by presentation format).
Note. Error bars represent standard error of measurement.

$p < .001$, partial $\eta^2 = .63$, and the column condition, $F(2, 24) = 20.61$, $p < .001$, partial $\eta^2 = .63$. There was a trend toward an interaction between age and condition, such that college students were more sure in the column condition, but third and sixth graders were more sure in the pairwise condition, $F(2, 95) = 2.37$, $p = .099$, partial $\eta^2 = .05$. The mean sureness ratings are presented in Figure 4 and indicate a clear pattern of decreased sureness ratings with increased variability in college students across both conditions.

As in Experiment 1, we calculated t tests for each set of data participants saw. In the pairwise condition, the t values ranged from 0 to 3.98. For this condition, we again found evidence of a strong association in college students between average sureness ratings and t statistics ($r = .924$, $p < .001$). Third graders exhibited a similar pattern ($r = .685$, $p = .042$). Sixth graders did not demonstrate a clear association ($r = .484$, $p = .177$). In the column condition, a different pattern emerged. These t values also ranged from 0 to 3.98, but only the college students showed any evidence of using this information systematically. College students' confidence in the difference between groups was very closely matched to actual t statistics ($r = .906$, $p = .001$). In contrast, third and sixth graders' ratings showed no evidence of a positive association between confidence and t statistics (third grade: $r = -.085$, $p > .5$; sixth grade: $r = .035$, $p > .5$).

Reasons offered. To classify explanations for reasoning, we began with the same codes as in Experiment 1. Two coders independently coded each set of explanations. The overall agreement level was 86.3%, and all discrepancies were resolved through discussion.

The pattern of explanations offered was similar to that in Experiment 1, with large grade effects. However, there were some key differences between the patterns of responses offered in Experiments 1 and 2. The frequencies of participants mentioning each explanation are summarized in Tables 4 and 5. The incremental data presentation led to more frequent mentions of data characteristics such as sample size (e.g., "The engineer should test again to get more data"; "Don't need to test again because we already have a lot of data"), suggesting that this format increased the salience of this characteristic. In Experiment 1, only 10% of third graders and 27% of sixth graders mentioned sample size. However, when directly comparing data sets with increasing data, 40% of third graders and 71% of sixth graders mentioned sample size in the pairwise condition, whereas 33% and 47%, respectively, did so in the column condition. Nearly all college students mentioned sample size. In discussing variability, mention of reversals or lack thereof was more common in the pairwise condition across all age groups. Mention of spread was similarly uncommon for third and sixth graders in the column condition as in Experiment 1 (8% and 13% as compared to 3% and 15%, respectively), and not a single third or sixth grader referred to spread in the pairwise condition.

Participants were asked to reason about all the data available in each scenario. However, as noted earlier, we found some participants explicitly stating that they only were referring to a single pair of data points in drawing conclusions. In Experiment 1, only 23% of third graders and 1 sixth grader ever clearly referred to a subset of the data in justifying answers. In

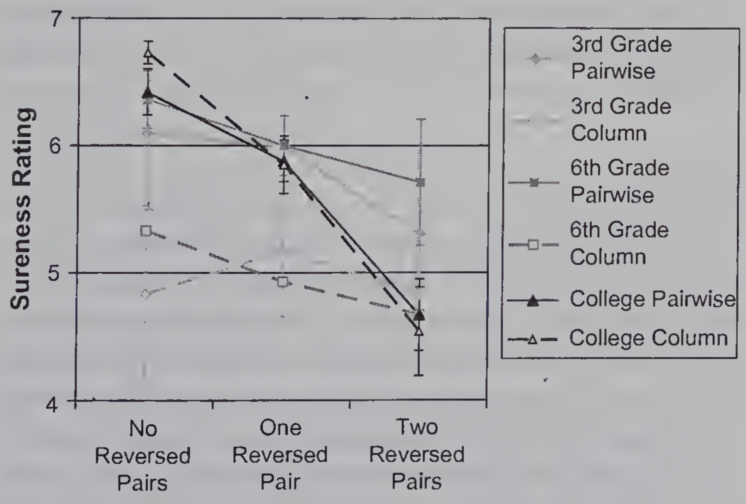


Figure 4. Experiment 2: Participants' sureness ratings at each grade with six pairs of data, by level of between-group variability in the data and presentation format.
Note. Error bars represent standard error of measurement.

Table 4
Experiment 2: Percentage of Participants at Each Age Who Mentioned Each Justification for Their Sureness, in the Pairwise Condition

	Third grade	Sixth grade	College
Data responses			
Sample size**	40	71	100
Reversed pairs**	30	57	83
No reversed pairs*	30	64	83
Spread of data	0	0	21
Overall trend in data	100	93	100
Single pair "trend"**	80	43	13
Magnitude of differences**	20	64	88
Magnitude of single pairs*	30	57	17
Inference**	30	43	88
Mechanism responses			
Ball property	20	36	21
Robot/athlete property*	40	57	13
Environment	0	14	8

Note. Age differences: $^*p < .05$. $^{**}p < .01$.

Table 5

Experiment 2: Percentage of Participants at Each Age Who Mentioned Each Justification for Their Sureness, in the Column Condition

	Third grade	Sixth grade	College
Data responses			
Sample size**	33	47	92
Reversed pairs	17	20	46
No reversed pairs**	8	20	69
Spread of data	8	13	35
Overall trend in data	83	87	100
Single pair "trend"	58	53	31
Magnitude of differences**	17	27	85
Magnitude of single pairs	58	47	38
Inference*	8	33	58
Unequal amount of data	50	53	73
Mechanism responses			
Ball property	8	27	12
Robot/athlete property	25	20	19
Environment	0	7	15

Note. Age differences: $*p < .05$. $**p < .01$.

contrast, the incremental presentation formats in Experiment 2 led many more participants to reason this way. As shown in Tables 4 and 5, many participants in all age groups referred to a pattern or trend in a single pair of data points or to the magnitude of the difference between the two numbers of a single pair in justifying their conclusions (e.g., after seeing the final two numbers presented in the column condition, referring to only these two numbers: "52 is far away from 40"; similar patterns occurred after seeing the last pair in the pairwise condition). In fact, the only significant age difference in Experiment 2 on these issues was in the pairwise condition, where 80% of third graders, 43% of sixth graders, and 13% of college students referred to a "pattern" or "trend" based on a single pair of data to justify their responses, $\chi^2(2, N = 48) = 14.54, p < .001$.

We also expanded the codes used in Experiment 1 to include more codes for data-based responses. In the column condition, it was common for participants to justify their request for more data with a reference to the fact that the columns were unequal; thus, it was common for participants to say that they would like more data in Column B because there was currently not as much data as in Column A (e.g., "There's four more numbers on A and there still needs to be four more on B so it will be even," "You need even amounts of data"). Therefore, we noted whenever participants referred to unequal sample sizes. In addition, because we asked participants to assess whether they wanted more data, many participants used the data to make inferences about whether more data were needed. We added the code of inference to account for answers such as "It's pretty clear that they're different" and "More data will confirm it one way or another."

In addition to discussions of data, we also explored whether participants offered causal mechanistic explanations as justifications for their conclusions. As in Experiment 1, we coded explanations for a mention of characteristics of the robot, characteristics of the ball, and characteristics of the environment. Overall, a little more than one third of the participants mentioned at least one mechanistic explanation (e.g., "Because while the robot threw the first two balls, he may have lost some power"; "There might be damage or air leakage in the ball"), with no significant age differences (36% of third graders, 45% of sixth graders, and 34% of college students mentioned such an explanation).

Testing again. In Experiment 2, participants were asked if they thought the engineers should test the balls again. If they said yes, they were asked whether the engineers should test Ball A, Ball B, or both. With only one data pair, participants asked for more data 93.4% of the time; with two data pairs, participants asked for more data 90.1% of the time; with four data pairs, participants asked for more data 71.6% of the time; and with six data pairs, participants asked for data 40.9% of the time. The grade breakdown showed an interesting pattern after all six data pairs were presented. Third graders wanted to test for more data about 60% of the time across between-group variability levels (zero, one, or two overlaps). Sixth graders wanted to test for more data about 30% of the time across levels. College students, however, were most likely to differentiate among the data sets. Thus, although only 14% of the college students asked for more data when there was little between-group variability, 40% asked for more data when there was one overlapping pair and 56% asked for more data when there were two overlapping pairs. See Figure 5 for details.

Discussion

Experiment 2 was conducted to determine if the relative salience of various data characteristics could be manipulated by varying presentation format. The results suggest that presentation did not change the pattern of explicit reasoning shown in Experiment 1 for college students. The majority of college students clearly changed their ratings systematically with sample size. There were no differences in ratings of sample size based on whether the data were presented pairwise or in columns, indicating that the processing of each format was comparable for sample

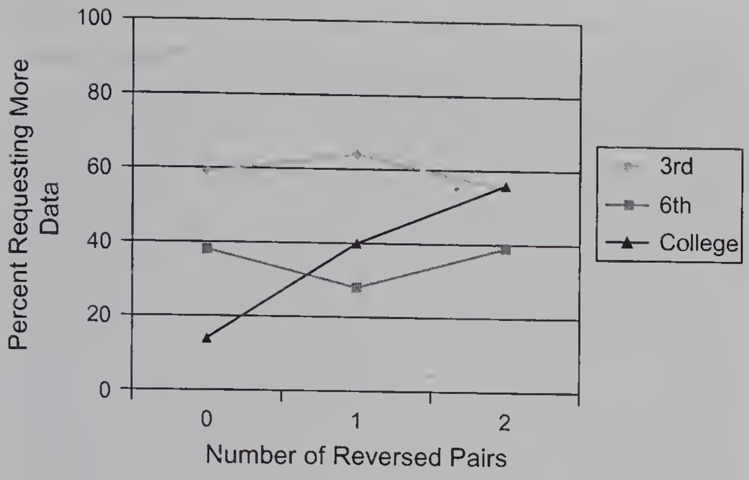


Figure 5. Experiment 2: Percentage of participants who requested further testing after six trials, based on between-group variability. *Note.* The data are broken down by age.

size ratings. However, third and sixth graders did not systematically use sample size when viewing data in these formats. This finding suggests that presenting data in this incremental format may in fact have made it somewhat more difficult for children to systematically use sample size when data sets they reasoned about were being updated.

However, the patterns of explicitly discussing sample size differed from Experiment 1. In Experiment 2, far more participants in the third and sixth grades talked about sample size as affecting their judgment. The presentation format increased the salience of this characteristic enough for many children to notice that sample size might play an important role. At the same time, incremental presentation of data led to more frequent references to trends in single pairs of data or subsets of the data instead of the whole pattern, though this trend was more pronounced for the younger participants. Because the task involved adding only one or two data points with each new trial, it is not surprising that this feature encouraged more participants to focus solely on a subset of the data rather than the whole, particularly in the pairwise condition.

Assessments of variability were also affected by data presentation format. College students demonstrated clear patterns of decreased sureness when there was more variability in the data, whereas the responses of younger participants were not as systematic. In addition, the format changes led to some differences in the explicit discussion of variability-related issues. In particular, participants appeared to be more likely to mention a lack of reversed pairs in the paired condition than in the column condition or in Experiment 1. Mention of spread of the data did not vary significantly by age, but mention of reversed pairs (or lack thereof) generally did. These findings suggest that children have a nascent understanding of variability but one that is not yet fully formed and applied equally across contexts.

General Discussion

The results of our study demonstrate that (a) children and adults attended to sample size and between-group variability when drawing conclusions from data; (b) there were significant age-related increases in the detection, use, and explicit awareness of individual data characteristics; and (c) the data presentation format affected the use of these characteristics. This research is the first to specifically examine the use of data characteristics outside knowledge-rich contexts and the ways these characteristics affect reasoning about data.

The framing in Experiment 1 allowed us to present two conditions (athlete, robot) that provided two different sources of error variance (e.g., human variation and robot variation, with the expectation that robots were likely to be perceived as less variable than humans). Although there were more mechanism-based explanations in the athlete condition than in the robot condition, there were few differences in how the data characteristics of sample size and variability were interpreted, suggesting that these characteristics were a source of information separate from domain knowledge. The fact that nearly every participant explicitly mentioned at least one data characteristic in justifying conclusions indicates an awareness of the importance of the numerical values. At the same time, Experiment 2 demonstrated that incremental data presentation led to more explicit reasoning about different data characteristics. These results suggest that participants may have difficulty reassessing sets of data with incremental additions to the data, even as these incremental additions make awareness of sample size more explicit. This pattern may indicate that initial values are disproportionately weighted relative to later values, though additional research is needed to evaluate this possibility. The most salient difference between presentation formats in Experiment 2 was in reasoning about between-group variability. In the column condition, participants still tended to use the characteristic but were slightly less sure of their conclusions, suggesting that the break from a focus on pairs often made them reason about the data set differently. However, when participants failed to consider the whole data set, they most often appear to have looked only at the most recently presented data and commented on a direct comparison on only this subset of the data.

The results demonstrate that children's data evaluations can be influenced by the characteristics of the data when comparing sets of data. But through what processes are reasoners making these comparisons? Though we did not specifically test this model, one possibility is that a mechanism for magnitude comparison and an explicit evaluation process may work in concert to produce the data interpretation described previously. Case and Okamoto (1996) provided evidence for a developmental model in which implicit processing mechanisms provide the foundation for children's learning in various domains such as number (i.e., central conceptual structures). Initially, development is highly constrained by processing limits that in turn constrain the number of dimensions a child can process, but as space and knowledge increase, children can attend to (and simultaneously process) multiple dimensions. Knowledge becomes increasingly available for explicit processing and can be extended to other areas (e.g., knowledge of living things). This model is one of many that suggest implicit processing mechanisms working in concert with gradually developing, explicit processes to account for developmental change within complex domains (e.g., Feigenson et al., 2004).

Our participants may have compared sets using approximate means and variances derived from the data. Though there is no definitive account of numerical estimation, one well-documented explanation suggests that number is represented by analog magnitudes that contain a proportion of error (as opposed to exact quantities; Gallistel & Gelman, 2000). There is evidence that estimating ability develops slowly and is difficult for many children and even adults (for discussion, see Siegler & Booth, 2005). Further evidence from recent research with adults suggests that estimating magnitude may be a general-purpose mechanism, showing similar patterns across varied information sources (Barth, Kanwisher, & Spelke, 2003). Although these studies presented data sets in a one-time comparison (not a series of data points to be combined later), our results suggest that the same representational mechanism used to represent and compare single quantities may represent and compare mean values. Obrecht, Chapman, and Gelman (2007) provided adults with a series of product ratings and found that participants were more confident in differences as mean differences increased and variance decreased, as in a *t* test.

If the process of mean computation is relatively automatic, then differentiating individual dimensions may require the acquisition of strategies that allow their identification and representation. Processing capacity (e.g., space, speed) likely plays a large role in the age-related performance differences (Gathercole et al., 2004; Kail, 2007). The transformation from presentation format (e.g., a set of values) into a different representational format (e.g., rough mean with error variance) requires sufficient space to implement these operations (Halford et al., 2007). In fact, this may be why children attended to sample size with simultaneous presentation (as in Experiment 1) but not with incremental presentation (as in Experiment 2). Research on processing capacity (e.g., Gathercole et al., 2004) indicates significant differences in working memory capacity between third and sixth graders and may help explain the age-related differences in attending to multiple data dimensions.

In addition, domain knowledge is likely necessary to identify individual data characteristics. Domain knowledge provides information about which dimensions may be associated with differences in behavior or performance. For example, if a professional golfer drives only 100 yards, most golf fans would ascribe this outcome to factors other than ability (e.g., wind, losing balance, losing concentration), whereas the same drive for a novice golfer might be ascribed to ability. In the present experiments, the domain knowledge was limited and was therefore only explicitly used some of the time in justifying conclusions. Setting the same data in a context in which participants have very strong prior theories may have led to different results and much greater reliance on this background knowledge (e.g., if we told participants Golfer A was a professional golfer and Golfer B was a complete beginner).

Related to domain knowledge is knowledge about using data themselves, particularly in understanding the relationship between a sample and a population. Increasing sample size increases the approximation to the population under study. Younger children in the present experiments did not appear to treat data as samples, as there was little effect of sample size. Paradoxically, third graders were often more certain with a single observation than with multiple observations. Perhaps, a source of young children's difficulty is an assumption that one observation is indicative of a population.

The results of this study suggest a prominent role for data in children's reasoning. Absent strong background knowledge, children make use of data as a source of information. Future research should more closely examine the influence of processing factors such as capacity and speed, in addition to varied representations. Future research should also examine how children's use of data (and data characteristics) might be useful for acquiring knowledge in novel domains.

Conclusions

In sum, these findings demonstrate that children and adults attend to the number and variability of observations when interpreting data. The results also demonstrate that the use of these characteristics improves with age and becomes more explicitly accessible. Response patterns are consistent with an intuitive *t* test, in which reasoners represent and compare an approximate mean that includes information on the variability of observations. With experience, children better identify the individual dimensions (e.g., variability), attend to finer distinctions of these dimensions, and increase their explicit access to this information. Although children's background knowledge is an important factor in interpreting data, our results indicate that the characteristics of the data themselves play a role in how children make sense of observations.

References

Barth, H., Kanwisher, N., & Spelke, E. (2003). The construction of large number representations in adults. *Cognition, 3,* 201–221.

Case, R., & Okamoto, Y. (1996). The role of central conceptual structures in the development of children's thought. *Monographs of the Society for Research in Child Development, 61*(1–2).

Chinn, C. A., & Malhotra, B. A. (2002). Children's responses to anomalous scientific data: How is conceptual change impeded? *Journal of Educational Psychology, 94,* 327–343.

Echevarria, M. (2003). Anomalies as a catalyst for middle school students' knowledge construction and scientific reasoning during science inquiry. *Journal of Educational Psychology, 95,* 357–374.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences, 8,* 307–314.

Gallistel, C. R., & Gelman, R. (2000). Non-verbal numerical cognition: From reals to integers. *Trends in Cognitive Sciences, 2,* 59–65.

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology, 40,* 177–190.

Gutheil, G., & Gelman, S. (1997). Children's use of sample size and diversity information within basic-level categories. *Journal of Experimental Child Psychology, 64,* 159–174.

Halford, G. S., Cowan, N., & Andrews, G. (2007). Separating cognitive capacity from knowledge: A new hypothesis. *Trends in Cognitive Sciences, 11,* 236–242.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence.* New York: Basic Books.

Jacobs, J. E., & Narloch, R. H. (2001). Children's use of sample size and variability to make social inferences. *Applied Developmental Psychology, 22,* 311–331.

Kail, R. V. (2007). Longitudinal evidence that increases in processing speed and working memory enhance children's reasoning. *Psychological Science, 18,* 312–313.

Klaczynski, P. A., & Aneja, A. (2002). Development of quantitative reasoning and gender biases. *Developmental Psychology, 38,* 208–221.

Klein, J. S., & Bisanz, J. (2000). Preschoolers doing arithmetic: The concepts are willing but the working memory is weak. *Canadian Journal of Experimental Psychology, 54,* 105–116.

Koerber, S., Sodian, B., Thoermer, C., & Nett, U. (2005). Scientific reasoning in young children: Preschoolers' ability to evaluate covariation evidence. *Swiss Journal of Psychology, 64,* 141–152.

Koslowski, B. (1996). *Theory and evidence: The development of scientific reasoning.* Cambridge, MA: MIT Press.

Koslowski, B., Okagaki, L., Lorenz, C., & Umbach, D. (1989). When covariation is not enough: The role of causal mechanism, sampling method, and sample size in causal reasoning. *Child Development, 60,* 1316–1327.

Kuhn, D., Amsel, E., & O'Loughlin, M. (1988). *The development of scientific thinking skills.* Orlando, FL: Academic Press.

Kuhn, D., Garcia-Mila, M., Zohar, A., & Andersen, C. (1995). Strategies of knowledge acquisition. *Monographs of the Society for Research in Child Development, 60*(4, Serial No. 245).

Lubben, F., & Millar, R. (1996). Children's ideas about the reliability of experimental data. *International Journal of Science Education, 18,* 955–968.

Masnick, A. M., & Klahr, D. (2003). Error matters: An initial exploration of elementary school children's understanding of experimental error. *Journal of Cognition and Development, 4,* 67–98.

Metz, K. E. (2004). Children's understanding of scientific inquiry: Their conceptualization of uncertainty in investigations of their own design. *Cognition and Instruction, 22,* 219–290.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84,* 231–259.

Obrecht, N. A., Chapman, G. B., & Gelman, R. (2007). Intuitive t-tests: Lay use of statistical information. *Psychonomic Bulletin and Review, 14,* 1147–1152.

Piaget, J., & Inhelder, B. (1975/1951). *The origin of the idea of chance in children.* New York: W.W. Norton.

Rubinsten, O., Henik, A., Berger, A., & Shahar-Shalev, S. (2002). The development of internal representations of magnitude and their association with Arabic numerals. *Journal of Experimental Child Psychology, 81,* 74–92.

Schauble, L. (1996). The development of scientific reasoning in knowledge-rich contexts. *Developmental Psychology, 32,* 102–119.

Siegler, R. S., & Booth, J. L. (2005). Development of numerical estimation: A review. In J. I. D Campbell (Ed.), *Handbook of mathematical cognition.* New York: Psychology Press.

Shaklee, H., & Paszek, D. (1985). Covariation judgment: Systematic rule use in middle childhood. *Child Development, 56,* 1229–1240.

Temple, E., & Posner, M. (1998). Brain mechanisms of quantity are similar in 5-year-old children and adults. *Proceedings of the National Academy of Sciences USA, 95,* 7836–7841.

Zimmerman, C. (2000). The development of scientific reasoning skills. *Developmental Review, 20,* 99–149.

Zimmerman, C. (2007). The development of scientific thinking in elementary and middle school. *Developmental Review, 27,* 172–223.

Child Development, July/August 2008, Volume 79, Number 4, Pages 1049–1064

Intergenerational Transmission of Social Anxiety: The Role of Social Referencing Processes in Infancy

Lynne Murray, Marc de Rosnay, Joanna Pearson, Caroline Bergeron, Elizabeth Schofield, Melanie Royal-Lawson, and Peter J. Cooper
University of Reading

Responses to an unfamiliar adult were examined in infants of mothers with social phobia ($N = 79$) and infants of nonanxious comparison mothers ($N = 77$) at 10 and 14 months in a social referencing paradigm. On each occasion, a female stranger first interacted with the mother and then approached and interacted with the infant. Over time, infants of mothers with social phobia showed increasing avoidance of the stranger, particularly when they were behaviorally inhibited. In boys, maternal social phobia was associated with increasing fearful responses. Infant avoidance was predicted by expressed maternal anxiety and low levels of encouragement to interact with the stranger. The findings are discussed in relation to theories concerning the intergenerational transmission of social anxiety.

Social phobia is one of the most common of the anxiety disorders, with a lifetime prevalence of 7%–13% (Furmark, 2002). In its generalized form, social phobia is characterized by a marked fear and avoidance of most novel social situations where the individual may be subject to scrutiny by others (e.g., participating in small groups, initiating or maintaining conversations), and by the fear that, in such situations, the individual will behave in a foolish or embarrassing manner (*Diagnostic and Statistical Manual of Mental Disorders* [4th ed.; *DSM–IV*], American Psychiatric Association, 1994). The disorder is severe and disabling and, without treatment, it tends to run a chronic course (Solyom, Ledwidge, & Solyom, 1986). Social phobia aggregates in families, with children of affected parents being themselves at risk (Lieb et al., 2000; Mancini, Van Ameringen, Szatmari, Fugere, & Boyle, 1996). A number of processes are likely to be involved in the intergenerational transmission of social anxiety. Genetic mechanisms are likely to contribute to risk (Eley, 2001), with twin studies showing significant, albeit moderate, genetic influence (Kendler, Karkowski, & Prescott, 1999; Nelson et al., 2000). Some of the variance in social anxiety is accounted for by genetic factors that are unique to this kind of anxiety (Kendler, Myers, Prescott, & Neale, 2001), but it appears that most of the genetic influence involves risk for emotional problems in general, including anxiousness and depression (Andrews, 1996; Eley, 1999). Relatedly, the role of temperament has been considered important in the development of social anxiety, and, in particular, it has been suggested that the temperamental constellation of behavioral inhibition (Kagan, Reznick, & Snidman, 1987) may represent a direct developmental precursor of the disorder (Stemberger, Turner, Beidel, & Calhoun, 1995). In favor of such a link is the finding that behavioral inhibition, usually measured from late infancy, raises the risk of anxiety disorder, especially the social type (Biederman, Rosenbaum, Bolduc-Murphy, & Faraone, 1993; Hirshfeld, Rosenbaum, Biederman, & Bolduc, 1992; Turner, Beidel, & Wolff, 1996).

The importance of environmental influences has also been stressed (Rapee & Spence, 2004). Among these, parenting practices have been emphasized, although direct evidence of their influence is limited (McLeod, Wood, & Weisz, 2007; Wood, McLeod, Sigman, Hwang, & Chu, 2003). This is particularly so with regard to mother–infant interactions, where samples have been small and have often included a variety of disorders. Furthermore, although previous research has consistently highlighted a number

---

The research was funded by a program grant from the Medical Research Council (United Kingdom). We thank Sheila Summers, Sue Campbell, Paula Libberton, Barbara Kennedy, Monika Parkinson, Rachel Kelly, and Mehjabeen Siddiqui for assistance with recruitment of subjects and assessment of mothers and infants. We thank Jerome Kagan and Nancy Snidman for providing training in the assessment of behavioral inhibition and for reliability assessment and confirmation for the research assistants. We thank Stephen Soldz, Margaret Charles, and Dave Grayson for their invaluable advice on statistical matters. We thank Tom O'Connor and Cathy Creswell for comments on the manuscript.

Correspondence concerning this article should be addressed to Lynne Murray, Winnicott Research Unit, School of Psychology, University of Reading, 3 Earley Gate, Whiteknights, Reading, RG6 6AL, UK. Electronic mail may be sent to lynne.murray@reading.ac.uk.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7904-0016

of parenting styles as posing a risk (e.g., overcontrolling/overprotective and unaffectionate parenting and the modeling of fearful behavior), concerns have been expressed about the robustness of these associations. In particular, it has been argued that insufficient account has been taken of the direction of effects, that there has been a lack of specificity in the conceptualization of the parenting practices concerned, and that assessment frames have been too general to identify the characteristics of specific disorders (Wood et al., 2003). There is, therefore, a need for longitudinal studies of parents with social phobia, in which systematic observations of parenting are made of adequate samples in contexts designed to elucidate the core clinical difficulties associated with the disorder.

A notable feature of social phobia, particularly of the generalized form, is its early onset. This suggests that in seeking to understand the development of social phobia, it will be valuable to study processes occurring early on, including those in infancy. Two features of the normal infant developmental process can be hypothesized to have a particular bearing on the development of social anxiety. First, at around 7–9 months of age, infants normally begin to show wariness in relation to unfamiliar adults, with increasingly negative reactions occurring over the rest of the 1st year (Sroufe, 1977). The second relevant process is *social referencing* or social information gathering; this concerns the way in which infants acquire information from other people concerning the emotional significance of aspects of their environment and how particular patterns of infant responsiveness to the environment may be engendered and maintained. Thus, from around 10 months, infants begin to seek information from their mother, or other relevant adult, concerning the adult's emotional response to, and appraisal of, a referent (e.g., an unfamiliar object or unfamiliar person); infants are then likely to modify their own behavior in line with that response. A large volume of research has accumulated concerning the conditions under which social referencing processes are likely to occur (Feinman, Roberts, Hsieh, Sawyer, & Swanson, 1992). Although there has been some variability in results, a number of findings have emerged of particular relevance to the question of the intergenerational transmission of anxiety. Thus, the impact of the adult's response may be especially marked in situations where the infant already experiences some uncertainty about the referent (Gunnar & Stone, 1984), and negative adult responses, including fearful responses, may be experienced by the infant as particularly salient (Hornik, Risenhoover, & Gunnar, 1987; Moses, Baldwin, Rosicky, & Tidball, 2001). Furthermore, at least with regard to infant responses to unfamiliar persons, the mother may be a privileged source of information (Feiring, Lewis, & Starr, 1984).

Taken together, these findings suggest an important route whereby social anxiety may be transmitted from mother to infant. Thus, at a time when the normative development of stranger fear emerges, the infant's vigilance concerning their caretaker's responses to other people is likely to result in any socially anxious behavior on the part of the mother leading the infant to adopt similar patterns of interpersonal responsiveness.

A number of processes may be involved in the transmission of socially anxious behavior from mother to infant. The first concerns the infant's or child's *observation* of their parent's response to the environment, sometimes referred to in the social referencing literature as an *indirect effect* (e.g., Feinman & Lewis, 1983) and the ensuing *modeling* by the child of the adult's behavior and affect (e.g., Gerull & Rapee, 2002; Muris, Steerneman, Merckelbach, & Meesters, 1996). A different set of processes concerns the parent's *direct* communication to the child about the referent and the ensuing management of the child's response. For example, the adult may encourage the child's engagement with the referent, thereby promoting child mastery and autonomy (e.g., Hornik & Gunnar, 1988). Conversely, they may attempt to limit the child's exposure to the referent (e.g., by taking over control of the object or the interaction with the other person themselves), thereby reinforcing and maintaining any initial child apprehension (e.g., Nachmias, Gunnar, Mangelsdorf, Parritz, & Buss, 1996; Rubin, Hastings, Stewart, Henderson, & Chen, 1997).

Here, we report on the development of socially anxious behavior in infants of mothers with social phobia and infants of nonanxious comparison mothers in a prospective, longitudinal study using a social referencing paradigm in which the mother was required to interact with an unfamiliar adult, a situation that was likely to elucidate the expression of any maternal social anxiety. Thus, a stranger first conversed with the mother in full view of infant, thereby allowing the infant to observe any signs of maternal anxiety and the nature of the mother–stranger interaction. Then, the mother was invited to introduce the stranger to the infant, whereupon the stranger approached and interacted with the infant. This second episode provided the opportunity for the infant to observe continuing maternal anxiety and also elucidated maternal management strategies potentially relevant to the transmission of anxiety. A particular merit of the social referencing paradigm is that it is analogous to many real-world social encounters in which the infant has the opportunity of observing parental management of ambiguous situations before making their own response.

We conducted an initial assessment at 10 months, when infants already show wariness or fear of strangers, but social-information-gathering strategies are not yet established, although they are nascent, and we made a second assessment at 14 months. A notable feature of social referencing research findings is that effects have generally been of only moderate magnitude; this is partly a function of the considerable individual differences in infant behavior. In view of this, we focused on the development of socially anxious behavior *within* individual infants from 10 to 14 months. We hypothesized that at a time when infants are both experiencing fearful responses to strangers and are monitoring the behavior of their caregivers as part of the process of social referencing, those whose mothers had social phobia would, compared to infants of comparison mothers, become progressively more fearful and avoidant of contact with the stranger over the period between the two observations.

Although offspring of individuals with social phobia are, in general, at raised risk for the disorder themselves, only a proportion are affected (e.g., in the study of Mancini et al., 1996, it was 25%), and it is important to investigate factors that may render some offspring more vulnerable than others. Thus, in addition to examining the overall effects of maternal social phobia on the development of infant socially anxious responses to a stranger, we investigated the effects of two further potentially important influences: first, the infant temperamental constellation of behavioral inhibition and, second, specific aspects of maternal behavior that have been linked to the occurrence of child anxiety (Wood et al., 2003).

Infant behavioral inhibition, which has commonly been measured at 14 months or later (e.g., Calkins, Fox, & Marshall, 1996; Fox, Henderson, Rubin, Calkins, & Schmidt, 2001; Kagan et al., 1987), refers to a general tendency to be fearful and withdrawn in the face of novelty. In the present study, where we were primarily concerned to examine the change in infant *social* fear and avoidance from 10 to 14 months, the inclusion in the definition of behavioral inhibition of infant responses to social novelty (i.e., to unfamiliar people) as a predictor variable would have entailed a fundamental confound with our outcome measure. We therefore restricted our assessment of infant behavioral inhibition to nonsocial contexts that have been widely used in the literature (see the Procedure section). We predicted that infants who were identified as inhibited in nonsocial contexts would show a greater increase in social fear and, in particular, avoidance in their interactions with the stranger over the critical developmental period between 10 months (when social information gathering is nascent) and 14 months (when infants are clearly able to engage in social referencing). We further hypothesized such infants would be more vulnerable to the effects of maternal social phobia than infants who were not inhibited; hence, we expected that behaviorally inhibited infants of mothers with social phobia to become more fearful and avoidant with the stranger over this time period.

With regard to the effects of maternal behavior on the infant's response to the stranger, we expected mothers with social phobia, in general, to express high levels of anxiety with the stranger, to be less encouraging of their infant's interaction with the stranger, and to be overcontrolling or intrusive when attempting to manage their infant's interaction with the stranger. Given that it is exactly these parenting behaviors that have consistently been found to be associated with anxiety in older children (Gruener, Muris, & Merckelbach, 1999; McLeod et al., 2007; Muris et al., 1996; Wood et al., 2003), we predicted that their occurrence would be associated with an increase in infant fear and avoidance of the stranger over the period from 10 to 14 months. Furthermore, we predicted that it would be by virtue of these features of the behavior of mothers with social phobia that their infants would develop increasingly negative responses to unfamiliar people; thus, we expected that once account was taken of these maternal behaviors, any effects of maternal social phobia on infant responses to the stranger would cease to be significant, that is, that such maternal behavior would mediate any effects of maternal social phobia on infant social responsiveness.

Finally, we addressed two subsidiary questions concerned with the issue of whether some infants may be more vulnerable to the effects of maternal social phobia than others. First, although effects of infant sex in the social referencing literature have traditionally been small and equivocal (Feinman et al., 1992), more recent evidence suggests that compared to boys, girls may modify their behavior more strongly when their mother expresses fear toward a referent (Blackford & Walden, 1998; Gerull & Rapee, 2002). We therefore took account of the infant's sex in the analyses. Second, because it is common in individuals with social phobia for there to be other concurrent anxiety disorders (Schneier, Johnson, Hornig, & Weissman, 1992), we conducted exploratory analyses to determine whether the presence of additional disorders affected the impact of maternal social phobia on infant development.

Method

Sample

Seventy-nine mothers with social phobia and 77 nonaffected comparison mothers, together with their infants, comprised the sample for the current study. They formed part of a larger sample, recruited from the Royal Berkshire Hospital in Reading, United Kingdom. Reading is a large town in the southeast of England, with a population (144,000) broadly reflecting the demographic characteristics of the country. Recruitment proceeded as follows: Four thousand mothers attending routine 20-week antenatal scan clinics at the hospital completed an eight-item screening questionnaire for social phobia. Three hundred sixty-six high-scoring women were recontacted; of these, 107 (29.2%) refused to participate in the study and 47 (12.8%) were found to be ineligible (e.g., because of moving away). The remaining 212 women were interviewed within 1 month to confirm diagnosis. Of these, 105 (49.5%) did not meet *DSM–IV* diagnostic criteria; 107 women were identified with social phobia of the generalized type (Stein et al., 1998) and were recruited. Twelve of these women subsequently became ineligible (e.g., due to premature delivery, infant death). A random sample of 164 women with low screening scores was also approached for participation in a comparison group: Thirty-two (19.4%) refused to take part and 9 (5.5%) were ineligible. The remaining 123 were interviewed to assess their mental state: Twenty-one were excluded as they had anxiety disorders other than social phobia and 7 mothers became ineligible; 1 woman was diagnosed with social phobia and was therefore recruited into the index group, bringing the final total in this group to 96. The remaining 94 women were recruited into the comparison group on the basis of having no current or past social phobia or any other anxiety disorder. The study sample was representative of those screened. Mean scores on the social phobia screening questionnaires for those who declined to participate did not differ significantly from those of women who were recruited and were, for high-scoring women, 23.05 ($SD = 4.05$) and 22.74 ($SD = 4.55$), respectively, $U = 5261.5$, $p = 0.31$. Similarly, for the low-scoring women, scores for those declining to participate and for those who were recruited into the comparison group were 9.75 ($SD = 2.27$) and 10.64 ($SD = 2.63$), respectively, $U = 1{,}311.5$, $p = 0.1$.

Of the 96 women with social phobia, 43 (45.7%) had no other anxiety disorder, 30 (31.9%) had comorbid generalized anxiety disorder, 21 (22.8%) also had a specific phobia, 8 (8.5%) had agoraphobia, 4 (4.3%) had panic disorder, and 3 (3.2%) had obsessive-compulsive disorder; in addition, 5 (5.2%) were depressed. One (1.1%) of the comparison group mothers was also depressed. The two study groups were similar in terms of demographic characteristics. Thus, for the comparison and social phobia groups, respectively, mean maternal ages were 31.2 ($SD = 4.2$) and 30.4 ($SD = 5.1$) years, and in each, the great majority of the mothers were Caucasian (94.7% and 100%) and were married or cohabiting (95.7% and 92.7%). Few were of low socioeconomic status (7.4% and 16.7%). Fifty percent of the comparison infants were boys, as were 39.6% of the social phobia group infants; infant birth weight was also comparable in comparison and social phobia groups ($Ms = 3{,}505$ g and 3,520 g, $SDs = 477$ and 411, respectively). At the 10 and 14 month assessments, because of infant illness or extreme distress at the start of testing, assessments could not always be conducted. The participants for whom data were available at both time points—79 (82.3%) of the social phobia group and 77 (81.9%) of the comparison group—did not differ from those without on maternal questionnaire responses or demographic characteristics (all $ps > .1$), and infant age at each assessment did not differ between groups ($Ms = 45.0$ weeks and 44.6 weeks, $SDs = 2.3$ and 2.2, respectively, for the first assessment and $Ms = 60.8$ weeks and 61.2 weeks, $SDs = 2.9$ and 3.8, respectively, for the second assessment, for comparison and social phobia groups, respectively).

Procedure

Mothers and their infants attended a university research unit at 10 and 14 months, when we conducted the assessment of infant social referencing and social anxiety; at 14 months, we also assessed infant behavioral inhibition. On each occasion, mothers completed rating scales concerning their current level of social anxiety. For the social referencing experiment, a researcher first described the procedure to the mother. Because we wished to observe any naturally occurring differences between mothers with social phobia and comparison mothers, unlike social referencing experiments in which mothers receive detailed instructions and training in how to behave, we asked the mothers to behave as they normally would. Mother and infant were then settled in an experimental room and were filmed via a one-way mirror. The infant was seated in a high chair with restraints, and the mother was seated 5 ft away from the infant. A female researcher, unfamiliar to mother and infant, entered the room, and, in the first phase of the experiment, she engaged the mother in conversation for a period of 2 min, asking her about the infant and her experience as a mother. In the second phase, the stranger informed the mother that she was going to approach the infant and asked the mother first to introduce her. The stranger then called to the infant and made a graded approach to the point where she stood close, looking at the infant directly; she stretched out her arms and touched the infant's arm. This approach phase lasted approximately 1 min. In the final phase, also lasting 1 min, the researcher unstrapped the infant from the chair, picked the infant up, and held him or her, positioning herself so that the infant and mother could see each other throughout. The stranger then talked to the infant in a friendly manner while watching the infant's face. The mother was free to interact with the stranger and the infant throughout.

For the assessment of behavioral inhibition, infants were observed in a paradigm in which latencies to approach and fearful or distressed behaviors toward novel objects and situations were assessed (Kagan et al., 1987). The assessment devised by Kagan et al. (1987), and used throughout the literature, involves exposure to both nonsocial and social (including stranger approach) stimuli. As already noted, in the present study because our outcome measure was the infant's response to an unfamiliar stranger and we wished to examine behavioral inhibition as a predictor or moderator of infant social anxiety, it would have been circular to use a behavioral inhibition index that included responses to social stimuli. Thus, to address the question of whether a general temperamental tendency to withdraw from novel and potentially threatening situations underlay any infant social avoidance, we used only the nonsocial elements of the assessment of behavioral inhibition. First, we recorded the infant's response to a mechanical dinosaur that moved and "roared"; the infant was seated on the carpet in the observation room and the dinosaur was placed approximately 3 ft away for a period of 3 min. Next, the infant was observed for 3 min in an unfamiliar playroom, with the mother present, where a collection of novel toys was available.

Measures: Maternal Mental State

(a) Screening questionnaire: An eight-item questionnaire derived from the Social Interaction and Anxiety Scale and the Social Phobia Scale (SIAS and SPS; Mattick & Clarke, 1998) was used. Items were included on the basis of their high factor loadings on these measures. Scores ranged from 8 to 37. A pilot of the screen showed a cutoff score of 17 to identify the top scoring 10% of respondents. (b) Diagnostic interview: Mothers were interviewed by trained psychologists and mental health clinicians with the Structured Clinical Interview for *DSM–IV* Axis 1 disorders (First, Spitzer, & Williams, 1995) to assess the presence of anxiety and depression. All interviews were subsequently discussed with a senior clinical team to confirm diagnosis and assignment to the two study groups. (c) Maternal social anxiety symptoms at 10 and 14 months: The full versions of the SPS and the SIAS were administered.

Measures: Maternal and Infant Behavior in the Social Anxiety Paradigm (10 and 14 Months)

On the basis of the literature on the parenting behaviors of mothers with anxiety disorders and that on infant social referencing, the following maternal and infant behaviors were scored.

Maternal behavior during the mother–stranger interaction. (a) Maternal anxiety: Signs of anxiety were rated on two 5-point scales: The first comprised ratings of facial expression (e.g., grimaces, biting lip, anxious/fearful expression) and bodily behaviors (e.g., wringing hands, tense posture) and the second, the quality of speech (e.g., nervous laughter, rushed and agitated, or else highly constrained, speech). Scores for the two scales were then summed. (b) Maternal engagement with the stranger: The extent of maternal talkativeness was rated on a 3-point scale; maternal avoidance of eye contact was similarly rated. The latter score was reversed and summed with the talkativeness score to give an overall measure of engagement.

Maternal behavior during the stranger–infant interaction. (a) Maternal anxiety: The same ratings were used as above. (b) Maternal encouragement: This rated the mother's positive encouragement of the infant's interaction with the stranger. Maternal nods and smiles, positive facial expressions, looks, and eyebrow flashes directed to the infant as the stranger approached and interacted with him or her were rated on a 3-point scale. (c) Maternal overcontrol/overprotection: This rated, on a 5-point scale, the extent to which the mother intervened in the stranger–infant interaction, for example, by interrupting and taking over the stranger's attempts to engage with the infant.

Infant behavior during the mother–stranger interaction. Infants varied in their emotional state at the start of the experiment. At 10 months, this appeared to be principally in response to being placed and restrained in the infant seat, although heightened wariness in response to the new surroundings was also possible. For these reasons, we took into account the extent of 10-month initial negativity, before the stranger's approach, when subsequently evaluating infant response to the stranger. Initial negative emotional tone was rated using a 5-point scale. This comprised ratings of facial and vocal expressions of distress, as well as bodily movements such as squirming and back arching. To ensure that the infant was able to obtain visual information about the mother's response to the stranger, whether the infant watched the mother was also scored as present or absent.

Infant behavior during the stranger–infant interaction. (a) Positive engagement: This was coded as present or absent and concerned positive behaviors such as smiling or laughing and clearly positive vocalizations directed to the stranger. (b) Infant fear: This was rated on a 5-point scale and included facial expressions of wariness (excluding avoidance) and fear and postural behaviors such as freezing. (c) Avoidance of stranger: Ratings were made on a 5-point scale and concerned active attempts to avoid contact, either through gaze avoidance or physically withdrawing or turning away. We examined infant avoidance of the stranger in two distinct periods during the infant–stranger interaction: the *approach* phase (during which the stranger approached the infant and made minimal physical contact) and the *pick-up* phase (during which the stranger unstrapped the infant and picked him or her up).

Maternal and infant behaviors during the mother–stranger interaction were scored in two equal time blocks of 1 min. During the infant–stranger interaction, there were four scoring blocks: three in the approach phase (each of 20 s) and one in the pick-up phase (1 min). This increased coding reliability and also meant that behavioral continuity within the interaction periods could be assessed.

Videotapes were scored by two researchers who were blind to both maternal group and the results of the assessment of infant behavioral inhibition. Twenty tapes selected at random were scored by both researchers to establish reliability. For continuous measures of maternal and infant behavior, Kappas ranged from .65 to .80. For the binary variables of infant positive engagement with the stranger and infant look to mother, Kappas were 0.79 and 1.00, respectively.

Infant behavioral inhibition in nonsocial settings. Infant behavioral inhibition was scored as present or absent in each of the four observation periods (one 3-min period for the playroom with unfamiliar toys and three 1-min periods for the dinosaur). For the playroom, where distress and fear were very rarely observed, an inhibited response was defined as the infant's maintaining close proximity to the mother and failing to approach the toys for at least 2 min of the whole 3-min period. For the dinosaur, an inhibited response was defined as the presence of fearful or distressed behavior, reflected in failure to approach the dinosaur, fretfulness and crying, scored in each 1-min period. We used both the continuous score of 0–4 (being the sum of the single playroom observation plus the three dinosaur observation scores) and a categorical measure of high behavioral inhibition (> 2 on the 0–4 point continuous score, reflecting the 25% most inhibited infants). Videotapes were scored by two trained researchers who were unaware of maternal group and of the results of the social referencing experiment. Twenty tapes were scored by both researchers to establish reliability. Mean Kappas for the continuous measures were 0.89, and for the categorical measure Kappa was 1.0.

Results

Continuing Maternal Social Anxiety Symptoms

There was considerable stability in mothers' experience of anxiety symptoms between the two assessments; rank order correlations being, for the SPS, .87, and for the SIAS, .91. Relatedly, the level of social anxiety experienced by the two study groups continued to differ; thus, the mean scores on the SPS for social phobia and comparison mothers at 10 months were 22.56 (*SD* = 13.90) and 5.59 (*SD* = 5.53), respectively ($U = 458.5, p < .001$), and at 14 months, 22.26 (*SD* = 14.09) and 5.48 (*SD* = 7.08), respectively ($U = 540.5, p < .001$). On the SIAS at 10 months, mean scores for the index and comparison groups were 39.70 (*SD* = 11.49) and 12.46 (*SD* = 7.21), respectively ($U = 105.5$, $p < .001$), and at 14 months, they were 39.38 (*SD* = 12.44) and 11.64 (*SD* = 8.89), respectively ($U = 150.5, p < .001$). The SPS and SIAS means for the social phobia mothers are comparable to those of clinic samples (Coelho, 2004).

Behavioral Inhibition

A similar proportion of index and comparison infants was classified as behaviorally inhibited: 18 (22.8%) and 15 (19.5%), respectively, $\chi^2(1) = 0.26$, *ns*; mean scores on the continuous measure for the two groups were 1.24 (*SD* = 1.30) and 1.08 (*SD* = 1.34), respectively, $t(154) = -.77$, *ns*.

Ratings of Maternal and Infant Behaviors in the Social Referencing Procedure

Examination of the frequencies and distributions of the infant and maternal behaviors at 10 months showed the occurrence of maternal overcontrol to be very infrequent (only 12 [7.7%] mothers); therefore, this variable was not considered further. With regard to infant behaviors, the measure of positive engagement with the stranger was significantly negatively correlated with that of fearfulness ($r = -.43, p < .01$); given the theoretical importance of fear of unfamiliar persons for this study, we retained this latter measure and excluded that of positive engagement from further analysis. The ratings of infant fear showed acceptably low levels of intercorrelation with those of avoidance (range = −.03 to .14, all *ns*). Notably, infant fear was highly consistent across the four scoring blocks of the infant–stranger interaction (Cronbach's $\alpha = .79$). By contrast, infant avoidance was far less stable (Cronbach's $\alpha = .37$), principally due to a marked escalation in avoidance when the stranger picked the infant up, thus confirming the need to treat behavior in this final pick-up phase as distinct from that during the stranger's approach. Inspection of distributions confirmed that it was appropriate to treat all continuous maternal and infant variables in their original, untransformed form.

Before investigating the change in infant behavior toward the stranger over time, and its determinants, we examined maternal and infant behavior in the two groups at 10 months, and the continuity in maternal behavior over time (see Table 1).

For maternal behavior at 10 months, those with social phobia showed significantly more anxiety than those in the comparison group, both during their own interaction with the stranger, $F(1, 154) = 4.16, p < .05$, and during the infant–stranger interaction, $F(1, 154) = 9.75, p < .01$. Index mothers were also significantly less engaged with the stranger than the comparison group, $F(1, 144) = 5.37, p < .05$, and they were somewhat less encouraging of infant interactions with the stranger, $F(1, 154) = 3.26, p = .07$. This

Table 1

Group Means (SDs) and Ranges for Maternal and Infant Behaviors at 10 and 14 Months

	10 months		14 months	
	Comparison	Social phobia	Comparison	Social phobia
Maternal behavior	(*n* = 77)	(*n* = 79)	(*n* = 77)	(*n* = 79)
Mother–stranger interaction				
Anxiety				
M(SD)	2.42 (0.79)	2.66 (0.70)	2.29 (0.65)	2.51 (0.66)
Range	3.50	2.50	3.00	3.00
Engagement				
M(SD)	2.37 (0.43)	2.21 (0.43)	2.38 (0.44)	2.24 (0.46)
Range	1.50	2.00	1.50	2.00
Infant–stranger interaction				
Anxiety				
M (SD)	2.68 (0.77)	3.04 (0.68)	2.62 (0.71)	2.95 (0.67)
Range	3.50	3.50	3.25	3.00
Encouragement				
M(SD)	1.78 (0.65)	1.61 (0.49)	1.61 (0.61)	1.59 (0.49)
Range	2.00	2.00	2.00	1.75
Infant behavior	(*n* = 73)	(*n* = 75)	(*n* = 73)	(*n* = 75)
Approach avoidance				
M (SD)	1.81 (0.48)	1.89 (0.60)	2.15 (0.67)	2.20 (0.64)
Range	2.00	2.67	2.67	3.00
Pick-up avoidance				
M (SD)	3.38 (0.70)	3.13 (0.72)	3.15 (0.95)	3.52 (0.92)
Range	3.00	3.00	4.00	4.00
Fearfulness				
M (SD)	2.21 (0.83)	1.94 (0.74)	2.28 (1.05)	2.19 (0.86)
Range	4.00	3.00	4.00	4.00
Nonsocial BI				
M (SD)			1.08 (1.34)	1.24 (1.30)
Range	—	—	4	4

Note. BI = behavioral inhibition.

pattern of group differences was retained at 14 months, and the extent of change in maternal behavior over time did not differ between groups. In contrast to maternal behavior, infant behavior at 10 months in the social referencing paradigm did not differ between the index and the comparison groups, for avoidance during the stranger's approach, $F(1, 136) = 1.18$, *ns*; avoidance during the stranger pick-up $F(1, 142) \leq 1$, *ns*; or fearfulness, $F(1, 136) \leq 1$, *ns*.

Change in Infant Social Behavior With the Stranger

Initial data checking showed that there were eight statistical outliers (four comparison [5%] and four index [5%]) on the avoidance measures during the 10-month assessment (> 2.5 *SD* from the mean) who were excluded from the main analyses reported below. However, including these infants did not alter the overall pattern of findings (including significance). Means, standard deviations, and ranges for infant behaviors during the infant–stranger interaction are shown in Table 1. All infants looked at their mother at some point during the time when she interacted with the stranger and during the stranger–infant interaction.

To examine change within individual infants' behavior from 10 to 14 months (calculated following the analytic strategy outlined by Bock, 1985, p. 489, by subtracting the 10-month from the 14-month score), and the extent to which infant behavioral inhibition, maternal group (comparison vs. social phobia), and maternal behavior at 10 months accounted for this change, we adopted a two-stage approach. First, we examined the influence of infant behavioral inhibition, maternal group, and their interaction (Inhibition × Group) using hierarchical linear multiple regression analyses. All continuous variables were centered, and categorical variables were coded as dummy variables. For each analysis, infant sex and negative emotional tone at 10 months were entered at the first step (base model). At the second step (main effects model), behavioral inhibition and maternal group were entered. At the final step (full model), the interaction term, Behavioral Inhibition × Group, was entered. Table 2 shows the Pearson correlation coefficients between predictor and outcome variables. The results of the regression analyses are presented in Table 3. In keeping with the overarching conceptualization of behavioral inhibition (Calkins et al., 1996; Kagan et al., 1987), we present these analyses with behavioral inhibition classified as a binary variable. In fact, results were not materially altered when it was treated as a continuous measure.

For the second stage of our analyses, we examined further the interactions identified in the first stage and explored the possibility of additional interactions between predictor variables where appropriate. We also conducted *process analyses* to establish whether 10-month maternal behaviors mediated any significant relations between behavioral inhibition, group or Behavioral Inhibition × Group, and change in infant behavior with the stranger, using the four steps outlined by Baron and Kenny (1986) and Preacher and Hayes (2004). We focus on 10-month maternal behaviors because of the importance, within a developmental framework, of establishing a predictive association between maternal behavior and change in infant responses to the stranger. As a final check, however,

Table 2
Correlations Between Predictor Variables and Change in Infant Behavior With the Stranger (N = 148)

	Δ avoidance: stranger approach	Δ avoidance: stranger pick-up	Δ fearfulness
Infant behavior with the stranger from 10 to 14 months			
Sex (0 = *male*, 1 = *female*)	−.02	−.06	$.16^{\dagger}$
INET 10 months	−.06	−.18*	−.22**
BI (0–4)	.01	−.04	$.14^{\dagger}$
BI (uninhibited, inhibited)	.06	.02	.09
Group (1 = *comparison*, 1 = *SP*)	−.02	.27**	.09
10-month maternal behavior (mother–stranger interaction)			
Anxiety	$.14^{\dagger}$	.18*	.04
Engagement	−.03	−.12	−.13
10-month maternal behavior (infant–stranger interaction)			
Anxiety	.10	.23**	.03
Encouragement	−.27**	−.17*	−.05

Note. INET = initial negative emotional tone; SP = social phobia; BI = behavioral inhibition.
$^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$.

Table 3

*Multiple Regression Analyses Evaluating the Effects of Maternal Social Phobia and Infant Behavioral Inhibition on Change in Infant Response to the Stranger (*N *= 148)*

	Infant behavior with the stranger from 10 to 14 months								
	Δ avoidance: stranger approach[a]			Δ avoidance: stranger pick-up[b]			Δ fearfulness[c]		
Variable	*B*	*SE B*	β	*B*	*SE B*	β	*B*	*SE B*	β
Base model									
Sex	−.03	.13	−.02	−.12	.19	−.05	.36	.17	.17*
INET 10 months	−.04	.05	−.06	−.17	.08	−.18*	−.20	.07	−.23**
Main effects model									
Sex	−.03	.13	−.02	−.16	.19	−.08	.35	.17	.17*
INET 10 months	−.04	.06	−.06	−.14	.08	−.14†	−.20	.07	−.23**
BI	.12	.16	.06	.04	.22	.02	.25	.20	.10
Group	−.04	.13	−.03	.59	.19	.25**	.08	.17	.04
Full model									
Sex	.04	.13	.02	−.13	.19	−.06	.37	.18	.17*
INET 10 months	−.04	.05	−.06	−.13	.08	−.14†	−.20	.07	−.23**
BI	−.24	.23	−.13	−.22	.33	−.08	.18	.30	.07
Group	−.19	.15	−.12	.48	.21	.21*	.05	.20	.02
BI × Group	.66	.32	.27*	.50	.46	.14	.13	.42	.04

Note. INET =initial negative emotional tone; BI = behavioral inhibition.

[a] $R^2 = .00$ for base model, ΔR^2 for main effects model = .00, ΔR^2 for full model = .03.

[b] $R^2 = .04$ for base model, ΔR^2 for main effects model = .06, ΔR^2 for full model = .01.

[c] $R^2 = .08$ for base model, ΔR^2 for main effects model = .01, ΔR^2 for full model = .00.

†$p < .10$. *$p < .05$. **$p < .01$.

we also examined whether any potential effects of 10-month maternal behaviors were better accounted for by 14-month maternal behaviors.

Infant avoidance of the stranger in the approach phase. During the stranger approach, only the interaction term, Behavioral Inhibition × Group, significantly contributed to change in stranger approach avoidance from 10 to 14 months. This interaction is depicted in Figure 1a, which shows that behaviorally inhibited infants of mothers with social phobia (n = 18) had a greater increment in avoidance during the stranger approach than noninhibited infants of index mothers (n = 57; see right-hand side of Figure 1a); a post hoc t test of this mean difference (.41) was marginally significant (p = .05). By contrast, behaviorally inhibited infants of comparison mothers (n = 15) had the smallest increment in avoidance to the stranger during the approach phase; a post hoc t test of the mean difference (.47) between inhibited infants of index and comparison mothers approached significance ($p < .10$). No further interactions (e.g., Sex × Group) were significant. Exploratory analyses (e.g., see Table 2) indicated that maternal encouragement might mediate the effects of the interaction between infant behavioral inhibition and maternal social phobia on the change in the infant's avoidance of the stranger, that is, a *mediated moderation effect* (Baron & Kenny, 1986; Muller, Judd, & Yzerbyt, 2005). The first step of the mediational analysis, requiring the association between the initial variable (Behavioral Inhibition × Group) and the outcome (change in infant approach phase avoidance), was already established (see Table 3). For Step 2, a relation needs to be shown between the initial variable and the mediator (maternal encouragement). A further regression analysis was run that confirmed this association, with the interaction between infant behavioral inhibition and group being significantly related to maternal encouragement, β = .27, $p < .05$, once the other study variables had been accounted for. Step 3—which establishes whether the mediator affects the outcome variable when controlling for the initial variable—and Step 4—which establishes the extent of mediation—are estimated in the same equation. Thus, a final regression was run in which maternal encouragement was also included as a predictor of the change in stranger approach avoidance, along with the effects of Behavioral Inhibition × Group. The results are shown in Table 4(a). The overall model was significant, $F(6, 141) = 2.68$, $p < .05$, and explained 10% of the variance, the majority of which was uniquely explained by maternal encouragement.

Table 4(a) shows that maternal encouragement was significantly related to change in stranger approach

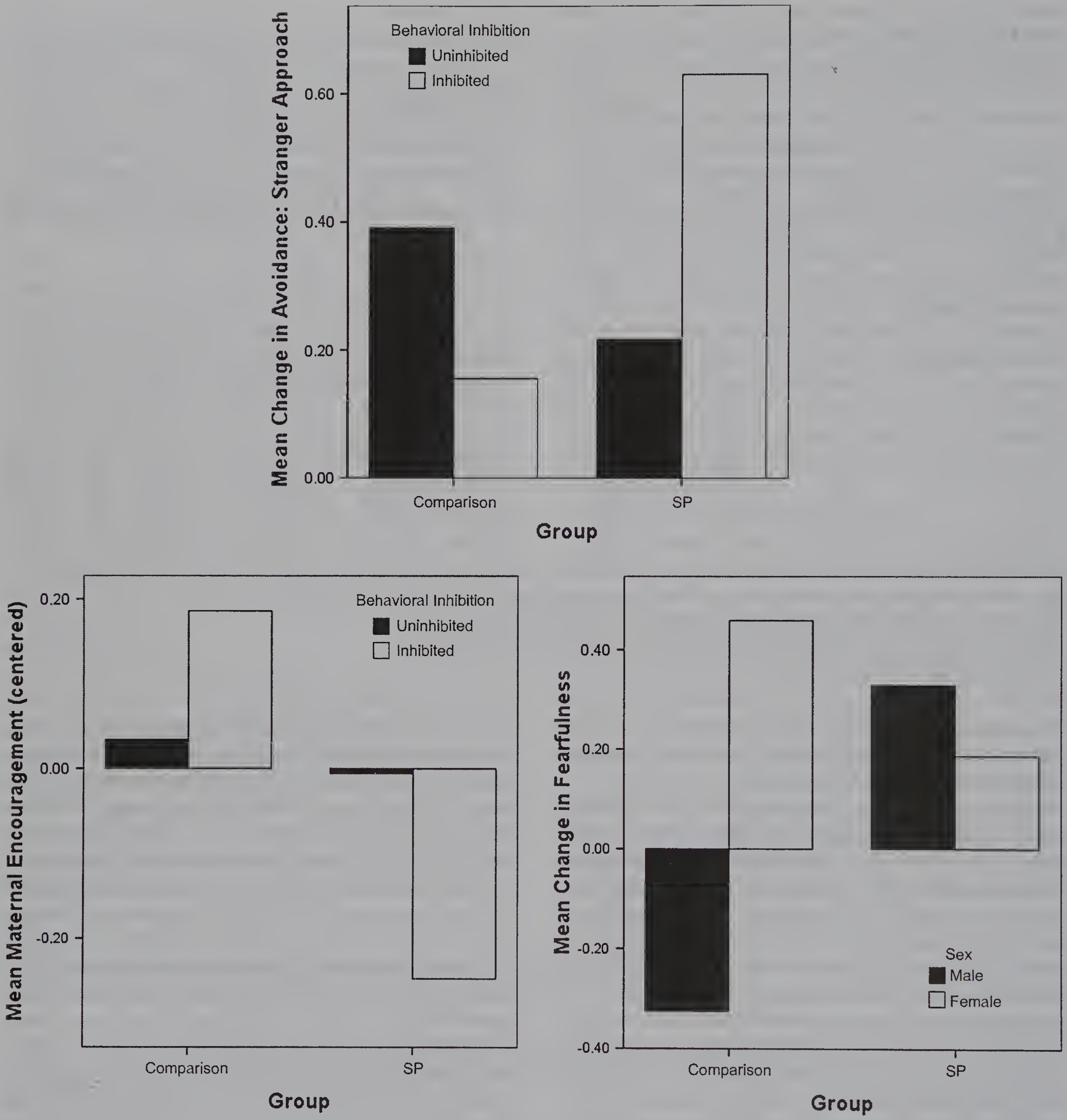


Figure 1. (a) Change in infant avoidance during the stranger approach by maternal group and infant behavioral inhibition. (b) Differences in maternal encouragement by group and infant behavioral inhibition. (c) Change in infant fearfulness by sex and group.

avoidance once the initial variable (Behavioral Inhibition × Group) and the other covariates were accounted for. Table 4(a) also shows that the relationship between Behavioral Inhibition × Group and change in stranger approach avoidance was no longer significant once maternal encouragement was included, indicating a mediating effect of this maternal behavior. To better understand the mediating influence of maternal encouragement, Figure 1b shows how maternal encouragement scores differed by group and behavioral inhibition (maternal encouragement scores were centered to have a mean of zero). It can be seen that index mothers of behaviorally inhibited infants were the *least* encouraging of all groups, and critically, comparison mothers of behaviorally inhibited infants were the *most* encouraging: A post hoc *t* test of the

Table 4

Multiple Regression Analyses Evaluating: (a) Maternal Encouragement as a Mediator of the Relation Between Behavioral Inhibition × Group and Change in Infant Avoidance of the Stranger During the Approach Phase and (b) Maternal Anxiety (MA) as a Mediator of the Relation Between Group and Change in Infant Avoidance of the Stranger During the Pick-Up Phase (N = 148)

Variable	*B*	*SE B*	β
(a)			
Sex (0=*male*, 1=*female*)	.00	.13	.00
INET 10 months	−.06	.05	−.09
Encouragement 10 months	−.38	.12	−.27**
BI (0 = *uninhibited*, 1 = *inhibited*)	−.16	.23	−.09
Group (0 = *comparison*, 1 = *SP*)	−.21	.15	−.13
BI × Group	.49	.32	.20
(b)			
	MA measured during mother – stranger interaction		
Sex (0 = *male*, 1 = *female*)	−.19	.18	−.08
INET 10 months	−.15	.08	−.15†
Anxiety	.24	.12	.16†
BI (0 = *uninhibited*, 1 = *inhibited*)	.03	.22	.01
Group (0 = *comparison*, 1 = *SP*)	.53	.19	.23**
	MA measured during infant – stranger interaction		
Sex (0 = *male*, 1 = *female*)	−.16	.18	−.07
INET 10 months	−.15	.08	−.15†
Anxiety	.29	.13	.18*
BI (0 = *uninhibited*, 1 = *inhibited*)	.02	.22	.01
Group (0 = *comparison*, 1 = *SP*)	.48	.19	.21*

Note. INET = initial negative emotional tone; BI = behavioral inhibition.
$^{\dagger}p < .10$. *$p < .05$. **$p < .01$.

mean difference (.43) was significant ($p < .05$). Controlling for 14-month maternal encouragement did not alter the overall pattern of these findings.

Infant avoidance of the stranger during the pick-up phase. Neither the principal interaction term of interest (i.e., Behavioral Inhibition × Group) nor other interaction terms (e.g., Sex × Group) significantly influenced change in stranger avoidance during the pick-up phase. Thus, subsequent process analyses were based on the main effects model, which had shown a significant effect of maternal group (see Table 3), whereby infants of mothers with social phobia became more avoidant between 10 and 14 months than infants of comparison mothers.

Exploratory analyses (see Table 2 and the Results section) indicated that maternal expressed anxiety—both during the mother – stranger and infant – stranger interactions—might mediate the relationship between group and change in stranger pick-up avoidance. We provide separate analyses for these two indices of maternal anxiety because they have, arguably, different implications for the mechanisms underpinning the transmission of social anxiety from mother to infant (i.e., indirect and direct transmission, respectively).

The main effects model (Table 3) satisfies Step 1 of the mediation analysis. For Step 2, two regression analyses were conducted—one each for maternal anxiety expressed during the mother – stranger and infant – stranger interactions—that examined the association between maternal anxiety and group. The first analysis, concerning anxiety expressed during the mother – stranger interaction at 10 months, suggested a marginal association with group, $\beta = .16$, $p = .05$, once the other study variables had been accounted for. The second analysis, concerning anxiety expressed during the infant – stranger interaction, suggested that once the other study variables had been accounted for, group was significantly related to maternal anxiety at 10 months, $\beta = .25$, $p < .01$. Because both indices of maternal anxiety satisfied the Step 2 criterion for mediation, two further regression analyses (Steps 3 and 4) were run that included maternal anxiety as a predictor of change in stranger pick-up avoidance as well as group. The results of these models are presented in Table 4(b).

The overall model including maternal anxiety during the mother – stranger interaction was significant, $F(5, 142) = 3.93, p < .01$, and explained 12% of the variance in the change in stranger pick-up avoidance. Maternal anxiety was marginally significant ($p = .05$) and, to some extent, satisfied the criterion for mediation, though it is notable that this mediation was marginal, and group still had a significant impact on change in stranger pick-up avoidance.

The overall model including maternal anxiety during the infant – stranger interaction was significant, $F(5, 142) = 4.23$, $p < .001$, and explained 13% of the variance in change in stranger pick-up avoidance. Maternal anxiety was significant ($p < .05$) and satisfied the criterion for partial mediation; group still had a significant impact on change in stranger pick-up avoidance, although there was a noticeable reduction in the strength of this relationship (see Table 3 for a comparison). Controlling for 14-month maternal anxiety during both the mother – stranger and the stranger – infant interactions did not alter the overall pattern of findings reported in Table 4.

Infant fearfulness of the stranger during the stranger – infant interaction. The first stage of the analyses revealed that neither infant behavioral inhibition nor group (nor their interaction) contributed significantly to change in infant fearfulness (see Table 3). However, the base model indicated that infant sex was significant, with girls becoming more fearful over time. Further investigation, however, showed an interaction between group and sex. Table 5 shows the full model containing initial infant emotional tone, sex, group, and Sex × Group; there were no other significant interactions. The overall model was significant, $F(4, 143) = 4.87, p < .001$, and explained 12% of the variance in change in infant fearfulness. Table 5 shows that all variables made a significant independent contribution. There were main effects for sex and group, which indicated that female infants and infants of index mothers became more fearful between 10 and 14 months. These main effects were qualified, however, by a significant Sex × Group interaction (depicted in Figure 1c): Among comparison mother–infant pairs, male infants became less fearful and female infants became more fearful between 10 and 14 months. Against this backdrop, it is notable that male, as well as female, infants of mothers with social phobia became more fearful from 10 to 14 months. Post hoc t tests confirm that the decrement in fearfulness for male infants of comparison mothers differed significantly from all other groups ($p < .05$) but that there were no other significant differences between the other combinations.

Effects of additional disorders. Finally, we investigated, in exploratory, post hoc analyses, whether the effects of maternal social phobia on infant social responsiveness to the stranger varied according to the presence or absence of other disorders. Thus, the change in infant pick-up avoidance from 10 to 14 months was compared between the group of infants whose mothers had only social phobia and those whose mothers had social phobia plus another diagnosis: No difference was apparent.

Discussion

Offspring of mothers with social phobia are themselves at risk for the disorder (Mancini et al., 1996), but little is known about how early social anxiety develops. Parenting has been considered important, but previous studies of anxious adults and their children have been subject to methodological difficulties. The current study sought to address these problems: using a large, carefully assessed, index sample, all of whom had social phobia, together with a nonanxious comparison sample, we conducted a prospective study in which direct observations of parenting were made under conditions specifically designed to elucidate the core difficulties associated with the maternal disorder and at a time when infants are likely to be particularly susceptible to the impact of these difficulties. We also sought to address the question of whether some infants may be more at risk than others for the development of socially anxious behavior. We therefore took account of infant individual differences in the temperamental constellation of behavioral inhibition, as well as differences in infant exposure to parenting styles associated with the occurrence of child social anxiety.

A central finding of the current study was that infants of mothers with social phobia became progressively more avoidant of social interactions with an unfamiliar adult between 10 and 14 months, particularly during stressful periods of stranger–infant interactions (i.e., during the pick-up phase). Moreover, where infants of mothers with social phobia were behaviorally inhibited, this increase in avoidance extended to the less stressful condition when the stranger made a graded approach. Finally, we found that infant fearful responses to be a function of both maternal social phobia and the infant's sex. Whereas girls in both the study groups showed an increase in their fear of the stranger over time, a pattern consistent with other reports of normal samples (Blackford & Walden, 1998; Gerull & Rapee, 2002), in boys the increase in fearful responses occurred principally in infants of index mothers. The results for infants of mothers with social phobia alone were similar to those of infants whose mothers experienced another disorder in addition to social phobia.

The infant avoidance and fearfulness shown in relation to the stranger's approach and attempts to interact with them in the social referencing paradigm clearly share features in common with the avoidance and anxiety surrounding social contacts of older children and adults with social phobia, and the literature on infant behavioral inhibition suggests that such infant behaviors are predictive of later clinically significant social anxiety (e.g., Stemberger et al., 1995). Nevertheless, it will be important to follow up the current sample to confirm that the

Table 5
Summary of Multiple Regression Analysis Evaluating Predictors of Change in Infant Fearfulness (N = 148)

Variable	B	$SE\ B$	β
Sex (0 = *male*, 1 = *female*)	.77	.24	.37**
INET 10 months	−.18	.07	−.21*
Group (0 = *comparison*, 1 = *SP*)	.55	.25	.26*
Sex × Group	−.84	.34	−.36*

Note. INET = initial negative emotional tone.
$*p < .05$. $**p < .01$.

infant responses to the stranger we observed do indeed predict subsequent child disorder.

Examination of maternal behavior, both during her own interaction with the stranger and during the stranger's interaction with the infant, helped us clarify some of the processes contributing to the changes in infant social responsiveness. Thus, mothers with social phobia were, as expected, more likely than comparison mothers to show signs of anxiety throughout the experiment. In turn, infants of mothers who expressed such anxiety, particularly as the stranger attempted to interact with them, became more avoidant of the stranger in the pick-up phase over time. We also found that low levels of maternal encouragement to the infant to interact with the stranger at 10 months predicted an increase in infant avoidance during the stranger's approach. These findings demonstrate, therefore, how processes previously shown with normal populations to be important in the transmission of fearful responses from mother to child may also be implicated in the intergenerational transmission of social anxiety in a clinical sample. They also contribute to the corpus of knowledge on the social referencing process by examining effects over time: Thus, although the immediate impact on infant behavior of negative maternal affect in relation to stimuli that may arouse fearfulness is well documented (e.g., Hornik et al., 1987; Moses et al., 2001), in the current study we found an impact of maternal behavior on the infant some months later. Specifically, 10-month maternal behaviors predicted change in infant behavior with the stranger, over and above the effects of any 14-month maternal behavior.

A number of our findings require further comment. The first concerns the role of infant behavioral inhibition. A number of previous studies have reported a link between behavioral inhibition and infant or child social anxiety, but there is a concern that, because of the way in which inhibition is measured (i.e., the inclusion of inhibited responses to *social* stimuli), the two variables may have been confounded. Indeed, a carefully conducted analysis of social and nonsocial inhibition in toddlers by Kochanska (1991) showed the two forms to be largely independent. In view of these concerns, we limited our measure of behavioral inhibition to responses to nonsocial stimuli. In common with the study of mothers with panic disorder by Warren et al. (2003), we did not find an increase in the rate of behavioral inhibition in infants of mothers with social phobia, nor was behavioral inhibition assessed in nonsocial conditions *directly* related to infant social avoidance. Nevertheless, this profile of infant behavior *was* of relevance for the development of infant social responsiveness. Most notable was the fact that mothers with social phobia reacted differently from comparison mothers to this infant behavior. Whereas the comparison group responded by being highly encouraging of their inhibited infants' interaction with the stranger, socially phobic mothers of inhibited infants showed the lowest levels of encouragement. Other studies also suggest that parenting practices that promote child social anxiety are particularly likely to occur in anxious parents in the context of child inhibited, anxious, or negative behavior (Hirshfeld, Biederman, Brody, Faraone, & Rosenbaum, 1997; Manassis & Bradley, 1994; Rubin, Nelson, Hastings, & Asendorpf, 1999; Woodruff-Borden, Morrow, Bourland, & Cambron, 2002). A number of factors are likely to come into play in these associations; thus, anxious parents are prone to overestimate their child's vulnerability, anticipate catastrophic outcomes in potentially threatening situations, and respond by promoting avoidant solutions to challenges (Wood et al., 2003). Translating these findings to the current study, it may have been the case, therefore, that mothers with social phobia underestimated their behaviorally inhibited infants' capacity to cope with the stranger contact, felt that their infants would find it distressing to have to engage with the stranger, and as a consequence, sought to reduce the possibility of contact by not encouraging infant engagement.

It was also notable that the link between infant behavioral inhibition and social avoidance was a function of the mother's response to the infant's inhibition. Thus, behaviorally inhibited infants in the comparison group, whose mothers were highly encouraging, showed the least increase in avoidance of the stranger over time. By contrast, behaviorally inhibited infants of index mothers, faced with the lowest levels of encouragement, showed the greatest increase in avoidance.

Three further aspects of our findings require comment. First, we found that maternal behavior during the stranger's approach and interaction ("direct" phase) with the infant was somewhat more clearly associated with infant responsiveness than maternal behavior that the infant initially observed in the mother–stranger interaction ("indirect" phase). This may have been because the phase of the infant–stranger interaction was more stressful for the mothers (indeed, levels of maternal expressed anxiety were higher at this time). In general, social referencing research has focused on the impact of the earlier phase of the paradigm. The fact that we did not find a strong effect of maternal behavior in this condition may be because, in this phase, the naturally occurring behavior of the mothers with social phobia in our study, while distinguishable from that of comparison mothers, was more subtle than that staged by mothers in experimental studies. Typically, mothers' staged behavior in experimental studies has required careful training, and such behaviors have been found to be considerably clearer and more intense than responses generated under unconstrained conditions (Rosen, Adamson, & Bakeman, 1992). It should be noted, however, that the somewhat greater influence of maternal behavior expressed during the infant's interaction with the stranger in our study is consistent with the results of other studies that have compared the impact of these two phases of the social referencing paradigm (e.g., Feinman & Lewis, 1983; Hornik & Gunnar, 1988).

An increase in infant fearful behavior with the stranger was associated with maternal social phobia only in boys. This may have been because in female infants, the normative increase in fear swamped any effects of maternal mental state. Alternatively, male infants may simply be more vulnerable than females to maternal psychopathology, a phenomenon observed in relation to maternal postnatal depression (e.g., Murray, Kempton, Woolgar, & Hooper, 1993; Weinberg, Olson, Beeghly, & Tronick, 2006). In spite of the association between fearfulness in boys and maternal social phobia, we were unable to establish the mechanisms mediating this link, with none of the maternal behaviors we measured predicting the increase in fear.

Finally, contrary to some suggestions in the literature, we did not find high rates of overcontrolling behavior in mothers with social phobia. However, most studies linking overcontrolling parenting to anxiety in offspring have been based on anxious adults' retrospective reports of the parenting they received or anxious children interacting with their parents. Such studies do not, therefore, directly address the issue of the influence of *parental* anxiety on parenting behavior. In fact, studies of anxious parents and their children, whether involving direct observations (Moore, Whaley, & Sigman, 2004; Turner, Beidel, Robertson-Nay, & Tervo, 2003; Woodruff-Borden et al., 2002), or interview and questionnaire methods (Hirshfeld et al., 1997; Lieb et al., 2000; Rubin et al., 1999), have found little evidence for raised rates of overcontrol; rather, and consistent with our findings, observations show anxious mothers to behave in a withdrawn, disengaged manner.

Our findings have implications for the understanding of the process of intergenerational transmission of anxiety. One mechanism proposed to be important in the clinical literature is the child's modeling of his or her anxious parent's behavior in relation to feared situations (e.g., Gerull & Rapee, 2002; Muris et al., 1996; Rachman, 1977); another is the way in which the parent actively manages the child's behavior, and in particular, the positive benefits of providing support for child autonomy and mastery (Chorpita & Barlow, 1998; Rapee, 1997). Our findings provide support for both processes. Consistent with modeling is the fact that the expression of maternal anxiety was predictive of increasing infant avoidance of the stranger during the stressful pick-up phase. It cannot be determined from the present study whether this increase arose because of a mood contagion effect, whereby the infant becomes generally negative in the face of maternal anxiety and simply generalizes this negativity to whatever is in his or her immediate environment, or whether it is an effect of a social referencing process that is specific to the stranger. Second, our finding that maternal encouragement of the infant's engagement with the stranger showed consistent associations with infant avoidant responses accords with models of intergenerational transmission of anxiety that emphasize parental management.

A number of limitations applied to our study. First, we used a very specific observational frame to assess the transmission of social anxiety. Although this had the advantage that highly relevant processes could be elucidated, more general aspects of parenting were not considered. These include the nature of the mother–child attachment relationship, which previous research has found to mediate continuities between early temperamental characteristics and the development of later behavioral inhibition (Calkins & Fox, 1992; Warren, Huston, Egeland, & Sroufe, 1997). In addition, as the work of Ladd and colleagues has shown (e.g., Ladd & Pettit, 2002), there are important aspects of socialization practices that only indirectly involve the parents, such as whether the child attends day care and is regularly exposed to strangers. Future research should include consideration of these aspects of the child experience. Given the fact that our sample was composed mainly of Caucasian, two-parent, middle-class families, it also needs to be established whether our findings generalize to other populations. A further limitation was the fact that we did not consider the influence of fathers. Research into the effects of maternal depression on child development has shown how the father may act as a buffer against the adverse impact of maternal depression (Conrad & Hammen, 1989), and similar effects may apply in the context of maternal social phobia. Finally, we did not include consideration of genetic mechanisms in the transmission of social anxiety.

References

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Andrews, G. (1996). Comorbidity and the general neurotic syndrome. *British Journal of Psychiatry, 168*, 76–84.

Baron, R., & Kenny, D. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173–1182.

Biederman, J., Rosenbaum, J., Bolduc-Murphy, E., & Faraone, S. (1993). A 3-year follow-up of children with and without behavioral inhibition. *Journal of the American Academy of Child & Adolescent Psychiatry, 32*, 814–821.

Blackford, J. U., & Walden, T. A. (1998). Individual differences in social referencing. *Infant Behavior and Development, 21*, 89–102.

Bock, R. D. (1985). *Multivariate statistical methods in behavioral research*. Chicago: Scientific Software.

Calkins, S., & Fox, N. (1992). The relation among infant temperament, security of attachment and behavioral inhibition at twenty-four months. *Child Development, 63*, 1456–1472.

Calkins, S., Fox, N., & Marshall, T. (1996). Behavioral and physiological antecedents of inhibited and uninhibited behavior. *Child Development, 67*, 523–540.

Chorpita, B., & Barlow, D. (1998). The development of anxiety: The role of control in the early environment. *Psychological Bulletin, 124*, 3–21.

Coelho, H. (2004). *Generalised social phobia and generalised anxiety disorder: A phenomenological and family study of a non-clinical sample*. Ph.D. thesis, University of Reading, UK.

Conrad, M., & Hammen, C. (1989). Role of maternal depression in perceptions of child maladjustment. *Journal of Consulting & Clinical Psychology, 57*, 663–667.

Eley, T. (1999). Behavioral genetics as a tool for developmental psychology: Anxiety and depression in children and adolescents. *Clinical Child Psychology Review, 21*, 21–36.

Eley, T. C. (2001). Contributions of behavioural genetics research: Quantifying genetic, shared environmental and nonshared environmental influences. In M. Vasey & M. Dadds (Eds.), *The developmental psychopathology of anxiety* (pp. 45–59). New York: Oxford University Press.

Feinman, S., & Lewis, M. (1983). Social referencing at ten months: A second-order effect on infants' responses to strangers. *Child Development, 54*, 878–887.

Feinman, S., Roberts, D., Hsieh, K., Sawyer, D., & Swanson, D. (1992). A critical review of social referencing in infancy. In S. Feinman (Ed.), *Social referencing and the social construction of reality in infancy* (pp. 15–54). New York: Plenum Press.

Feiring, C., Lewis, M., & Starr, M. D. (1984). Indirect effects and infants' reaction to strangers. *Developmental Psychology, 20*, 485–491.

First, M. B., Spitzer, R. L., & Williams, J. B. W. (1995). *Structured Clinical Interview for DSM-IV Axis I*. Biometrics Research Department, New York State Psychiatric Institute.

Fox, N., Henderson, H., Rubin, K., Calkins, S., & Schmidt, L. (2001). Continuity and discontinuity of behavioral inhibition and exuberance: Psychophysiological and behavioral influences across the first four years. *Child Development, 72*, 1–21.

Furmark, T. (2002). Social phobia: Overview of community surveys. *Acta Psychiatrica Scandinavica, 105*, 84–93.

Gerull, F., & Rapee, R. (2002). Mother knows best: Effects of maternal modelling on the acquisition of fear and avoidance behaviour in toddlers. *Behaviour Research and Therapy, 40*, 279–287.

Gruener, K., Muris, P., & Merckelbach, H. (1999). The relationship between anxious rearing behaviours and anxiety disorders symptomatology in normal children. *Journal of Behaviour Therapy and Experimental Psychiatry, 30*, 27–35.

Gunnar, M. R., & Stone, C. (1984). The effects of positive maternal affect on infant responses to pleasant, ambiguous and fear-provoking toys. *Child Development, 55*, 1231–1236.

Hirshfeld, D., Biederman, J., Brody, L., Faraone, S., & Rosenbaum, J. (1997). Expressed emotion toward children with behavioral inhibition: Associations with maternal anxiety disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 36*, 910–917.

Hirshfeld, D., Rosenbaum, J., Biederman, J., & Bolduc, E. (1992). Stable behavioral inhibition and its association with anxiety disorder. *Journal of the American Academy of Child & Adolescent Psychiatry, 31*, 103–111.

Hornik, R., & Gunnar. M. R. (1988). A descriptive analysis of infant social referencing. *Child Development, 59*, 626–634.

Hornik, R., Risenhoover, N., & Gunnar, M. (1987). The effect of maternal positive, neutral and negative affective communications on infant responses to new toys. *Child Development, 58*, 937–944.

Kagan, J., Reznick, J., & Snidman, N. (1987). The physiology and psychology of behavioral inhibition in children. *Child Development, 58*, 1459–1473.

Kendler, K. S., Karkowski, L. M., & Prescott, C. A. (1999). Fears and phobias: Reliability and heritability. *Psychological Medicine, 29*, 539–553.

Kendler, K., Myers, J., Prescott, C., & Neale, M. (2001). The genetic epidemiology of irrational fears and phobias in men. *Archives of General Psychiatry, 58*, 257–265.

Kochanska, G. (1991). Patterns of inhibition to the unfamiliar in children of normal and affectively ill mothers. *Child Development, 62*, 250–263.

Ladd, G., & Pettit, G. (2002). Parenting and the development of children's peer relationships. In M. H. Bornstein (Ed.), *Handbook of parenting: Vol. 5. Practical issues in parenting* (pp. 269–310). Mahwah, NJ: Erlbaum.

Lieb, R., Wittchen, H., Hofler, M., Fuetsch, M., Stein, M., & Merikangas, K. (2000). Parental psychopathology, parenting styles, and the risk of social phobia in offspring: A prospective longitudinal community study. *Archives of General Psychiatry, 57*, 859–866.

Manassis, K., & Bradley, S. J. (1994). The development of childhood anxiety disorders: Towards an integrated model. *Journal of Applied Developmental Psychology, 15*, 345–366.

Mancini, C., Van Ameringen, M., Szatmari, P., Fugere, C., & Boyle, M. (1996). A high-risk pilot study of the children of adults with social phobia. *Journal of the American Academy of Child & Adolescent Psychiatry, 35*, 1511–1517.

Mattick, R., & Clarke, C. (1998). Development and validation of measures of social phobia scrutiny fears and social interaction anxiety. *Behaviour Research and Therapy, 36*, 443–453.

McLeod, B., Wood, J., & Weisz, J. (2007). Examining the association between parenting and childhood anxiety: A meta-analysis. *Clinical Psychology Review, 27*, 155–172

Moore, P., Whaley, S., & Sigman, M. (2004). Interactions between mothers and children: Impacts of maternal and child anxiety. *Journal of Abnormal Psychology, 113*, 471–476.

Moses, L., Baldwin, D., Rosicky, J., & Tidball, G. (2001). Evidence for referential understanding in the emotions domain at twelve and eighteen months. *Child Development, 72*, 718–735.

Muller, D., Judd C. M., & Yzerbyt, V. Y. (2005). When moderation in mediated and mediation is moderated. *Journal of Personality and Social Psychology, 89*, 852–863.

Muris, P., Steerneman, P., Merckelbach, H., & Meesters, C. (1996). The role of parental fearfulness and modeling in children's fear. *Behaviour Research and Therapy, 34*, 265–268.

Murray, L., Kempton, C., Woolgar, M., & Hooper, R. (1993). Depressed mothers' speech to their infants and its relation to infant gender and cognitive development. *Journal of Child Psychology and Psychiatry, 34*, 1083–1101.

Nachmias, M., Gunnar, M., Mangelsdorf, S., Parritz, R., & Buss, K. (1996). Behavioral inhibition and stress reactivity: The moderating role of attachment security. *Child Development, 67*, 508–522.

Nelson, E., Grant, J., Bucholz, K., Glowinski, A., Madden, P., Reich, W., et al. (2000). Social phobia in a population-based female adolescent twin sample: Co-morbidity and associated suicide-related symptoms. *Psychological Medicine, 30*, 797–804.

Preacher, K., & Hayes, A. (2004), SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers, 36*, 717–731.

Rachman, S. (1977). The conditioning theory of fear acquisition: A critical examination. *Behaviour Research and Therapy, 15*, 375–387.

Rapee, R. M. (1997). Potential role of childrearing practices in the development of anxiety and depression. *Clinical Psychology Review, 17*, 47–67.

Rapee, R. M., & Spence, S. H. (2004). The etiology of social phobia: Empirical evidence and an initial model. *Clinical Psychology Review, 24*, 737–767.

Rosen, W., Adamson, L., & Bakeman, R. (1992). An experimental investigation of infant social referencing: Mothers' messages and gender differences. *Developmental Psychology, 28*, 1172–1178.

Rubin, K. H., Hastings, P. D., Stewart, S. L., Henderson, H. A. M., & Chen, X. (1997). The consistency and concomitants of inhibition: Some of the children all of the time. *Child Development, 68*, 467–483.

Rubin, K., Nelson, L., Hastings, P., & Asendorpf, J. (1999). The transaction between parents' perceptions of their children's shyness and their parenting styles. *International Journal of Behavioral Development, 23*, 937–957.

Schneier, F., Johnson, C., Hornig, M., & Weissman, M. (1992). Social Phobia: Comorbidity and morbidity in an epidemiological sample. *Archives of General Psychiatry, 49*, 282–288

Solyom, L., Ledwidge, B., & Solyom, C. (1986) Delineating social phobia. *British Journal of Psychiatry, 149*, 464–470.

Sroufe, L. A. (1977). Wariness of strangers and the study of infant development. *Child Development, 48*, 731–746.

Stein, M., Chartier, M., Hazen, A., Kozak, M., Tancer, M., Lander, S., et al. (1998). A direct-interview family study of generalized social phobia. *American Journal of Psychiatry, 155*, 90–97.

Stemberger, R. T., Turner, S. M., Beidel, D. C., & Calhoun, K. S. (1995). Social phobia: An analysis of possible developmental factors. *Journal of Abnormal Psychology, 104*, 526–531.

Turner, S., Beidel, D., Robertson-Nay, R., & Tervo, K. (2003). Parenting behaviors in parents with anxiety disorders. *Behaviour Research and Therapy, 41*, 541–554.

Turner, S., Beidel, D., & Wolff, P. (1996). Is behavioral inhibition related to the anxiety disorders? *Clinical Psychology Review, 16*, 157–172.

Warren, S., Gunnar, M., Kagan, J., Anders, T., Simmens, S., Rones, M., et al. (2003) Maternal panic disorder: Infant temperament, neurophysiology and parenting behaviours. *Journal of Academic Child and Adolescent Psychiatry, 42*, 814–825.

Warren, S., Huston, M., Egeland, B., & Sroufe, A. (1997). Child and adolescent anxiety disorders and early attachment. *Journal of Academic Child and Adolescent Psychiatry, 36*, 637–644.

Weinberg, M. K., Olson, K. L., Beeghly, M., & Tronick, E. Z. (2006). Making up is hard to do, especially for mothers with high levels of depressive symptoms and their infant sons. *Journal of Child Psychology and Psychiatry, 47*, 670–683.

Wood, J. J., McLeod, B. D., Sigman, M., Hwang, W., Chu, B. C. (2003). Parenting and childhood anxiety: Theory, empirical findings, and future directions. *Journal of Child Psychology and Psychiatry, 44*, 134–151.

Woodruff-Borden, J., Morrow, C., Bourland, S., & Cambron, S. (2002). The behavior of anxious parents: Examining mechanisms of transmission of anxiety from parent to child. *Journal of Clinical Child and Adolescent Psychology, 31*, 364–74.

Child Development, July/August 2008, Volume 79, Number 4, Pages 1065–1085

Family Resources and Parenting Quality: Links to Children's Cognitive Development Across the First 3 Years

Julieta Lugo-Gil and Catherine S. Tamis-LeMonda
New York University

Reciprocal associations among measures of family resources, parenting quality, and child cognitive performance were investigated in an ethnically diverse, low-income sample of 2,089 children and families. Family resources and parenting quality uniquely contributed to children's cognitive performance at 14, 24, and 36 months, and parenting quality mediated the effects of family resources on children's performance at all ages. Parenting quality continued to relate to children's cognitive performance at 24 and 36 months after controlling for earlier measures of parenting quality, family resources, and child performance. Similarly, children's early cognitive performance related to later parenting quality above other measures in the model. Findings merge economic and developmental theories by highlighting reciprocal influences among children's performance, parenting, and family resources over time.

Economists and developmental psychologists have long been concerned with the factors that promote positive developmental outcomes in children (see Foster, 2002; Haveman & Wolfe, 1995, for reviews). However, the lenses of these disciplines differ in significant ways. Economists investigate the effects of parents' skills and monetary and time resources on their children's educational attainment, health, consumption, and ultimate wealth (Aiyagari, Greenwood, & Seshadri, 1999; Becker, 1964, 1991; Ermisch & Francesconi, 2000). In comparison, developmental psychologists emphasize social capital, especially parenting quality, as a core influence on children's development (Bornstein, 2002; Collins, Maccoby, Steinberg, Hetherington, & Bornstein, 2000; Landry, Smith, & Swank, 2006). To date, few studies have integrated economic and developmental perspectives (Guo & Harris, 2000) and few do rarely document

---

Julieta Lugo-Gil is now at Mathematica Policy Research Inc., Princeton, NJ

This work was conducted at New York University's Center for Research on Culture, Development, and Education (funded by the National Science Foundation, Grant BCS 021859). The findings reported here are based on research conducted as part of the national EHS Research and Evaluation Project funded by the Administration on Children, Youth and Families (ACYF), U.S. Department of Health and Human Services under contract 105-95-1936 to Mathematica Policy Research Inc., Princeton, NJ, and Columbia University's Center for Children and Families, Teachers College, in conjunction with the EHS Research Consortium. The Consortium consists of representatives from 17 programs participating in the evaluation, 15 local research teams, the evaluation contractors, and ACYF. Research institutions in the Consortium (and principal researchers) include ACYF (Rachel Chazan Cohen, Judith Jerald, Esther Kresh, Helen Raikes, and Louisa Tarullo); Catholic University of America (Michaela Farber, Lynn Milgram Mayer, Harriet Liebow, Christine Sabatino, Nancy Taylor, Elizabeth Timberlake, and Shavaun Wall); Columbia University (Lisa Berlin, Christy Brady-Smith, Jeanne Brooks-Gunn, and Allison Sidle Fuligni); Harvard University (Catherine Ayoub, Barbara Alexander Pan, and Catherine Snow); Iowa State University (Dee Draper, Gayle Luze, Susan McBride, and Carla Peterson); Mathematica Policy Research Inc. (Kimberly Boller, Ellen Eliason Kisker, John M. Love, Diane Paulsell, Christine Ross, Peter Schochet, Cheri Vogel, and Welmoet van Kammen); Medical University of South Carolina (Richard Faldowski, Gui-Young Hong, and Susan Pickrel); Michigan State University (Hiram Fitzgerald, Tom Reischl, and Rachel Schiffman); New York University (Mark Spellmann and C.S. T.); University of Arkansas (Robert Bradley, Mark Swanson, and Leanne Whiteside-Mansell); University of California, Los Angeles (Carollee Howes and Claire Hamilton); University of Colorado Health Sciences Center (Robert Emde, Jon Korfmacher, JoAnn Robinson, Paul Spicer, and Norman Watt); University of Kansas (Jane Atwater, Judith Carta, and Jean Ann Summers); University of Missouri-Columbia (Mark Fine, Jean Ispa, and Kathy Thornburg); University of Pittsburgh (Beth Green, Carol McAllister, and Robert McCall); University of Washington School of Education (Eduardo Armijo and Joseph Stowitschek); University of Washington School of Nursing (Kathryn Barnard and Susan Spieker); and Utah State University (Lisa Boyce, Gina Cook, Catherine Callow-Heusser, and Lori Roggman). We thank Eileen T. Rodriguez, Mark Spellmann, Gigliana Melzi, and Hirokazu Yoshikawa for feedback and support. We acknowledge seminar participants at New York University and The Urban Institute for helpful discussions and comments on previous versions of this article. We also thank three anonymous reviewers for helpful comments on earlier versions of this manuscript. Any opinions, findings, and conclusions or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Correspondence concerning this article should be addressed to Julieta Lugo-Gil, Mathematica Policy Research Inc., P.O. Box 2393, Princeton, NJ 08543, or to Catherine S. Tamis-LeMonda, Department of Applied Psychology, New York University, 239 Greene Street, 5th Floor, New York, NY 10003. Electronic mail may be sent to jlugo-gil@mathematica-mpr.com or to catherine.tamis-lemonda@nyu.edu.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7904-0017

the dynamics of parenting, economic resources, and children's abilities across early developmental periods. The present study addresses these gaps by focusing on reciprocal and unique influences among measures of parental economic resources, parenting behaviors, and children's cognitive performance within and across children's first 3 years.

Integrating Developmental and Economic Perspectives

Building on the traditions of Vygotsky (1978) and Bruner (1983), the developmental literature has long demonstrated the beneficial effects of parenting quality on children's language development, literacy, cognition, and school readiness (Bornstein, 2002; Leseman & de Jong, 1998; Storch & Whitehurst, 2001). Three features of parenting have been acknowledged to promote positive outcomes in young children: sensitivity, cognitive stimulation, and warmth (also termed positive regard or positive affect; Ainsworth, Blehar, Waters, & Wall, 1978; Baumrind, 1973; Bloom, 1993; Bornstein, 2002; Carpenter, Nagell, & Tomasello, 1998; Hart & Risley, 1995; Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1991; Landry, Smith, Swank, Assel, & Vellet, 2001; Maccoby & Martin, 1983; Matas, Ahrend, & Sroufe, 1978; Nelson, 1973, 1988; Pettit, Bates, & Dodge, 1997; Rohner, 1986; Skinner, 1986; Snow, 1986; Tamis-LeMonda, Bornstein, & Baumwell, 2001; Tamis-LeMonda, Shannon, Cabrera, & Lamb, 2004; Watson, 1985; Woodward & Markman, 1998). *Parenting sensitivity* refers to parents' attunement to their children's cues, emotions, interests, and capabilities in ways that balance children's needs for support with their needs for autonomy. *Cognitive stimulation* refers to parents' didactic efforts to enrich their children's cognitive and language development by engaging children in activities that promote learning and by offering language-rich environments to their children. *Parents' warmth* refers to parents' expressions of affection and respect toward their children and is thought to support skills for learning such as mastery, security, autonomy, and self-efficacy.

In contrast to developmentalists' focus on parenting quality, economists continue to build on the seminal work of Becker (1964) who emphasized the ways in which family resources (e.g., money and time) influence children's learning outcomes (e.g., Blau, 1999; Ermisch & Francesconi, 2000; Haveman & Wolfe, 1995; Hill & O'Neill, 1994; Levy & Duncan, 2000; T. W. Schultz, 1973; T. P. Schultz, 1993). Economists align with developmental psychologists in asserting that parents with fewer resources encounter numerous constraints in meeting their children's needs. However, economists rarely include direct observations of the quality of parent–child interactions in their assessments. Instead, economists conceptualize parenting quality as investments of time and money (Aiyagari et al., 1999; Becker, 1991; Davies & Zhang, 1995; Foster, 2002). Parents who spend more time and monetary resources on children's consumption, education, and care have children who achieve more in terms of cognitive development, educational achievement, and future earnings (e.g., Brown & Flinn, 2006; Hill & O'Neill, 1994; Moore, Evans, Brooks-Gunn, & Roth, 2001).

Although studies by developmentalists and economists have largely been pursued independently, there has been heightened awareness of the need to consider how parenting quality and economic resources *jointly* influence children's developmental outcomes. Studies have revealed that sustained poverty has significant negative effects on children's developmental outcomes and that the quality of parenting plays an important role in mediating these effects particularly at early ages (Bradley et al., 1989; Duncan, Brooks-Gunn, & Klebanov, 1994; Duncan, Yeung, Brooks-Gunn, & Smith, 1998; Garrett, Ng'andu, & Ferron, 1994; Mayer, 1997; National Institute of Child Health and Human Development [NICHD] Early Child Care Research Network, 2005; Yeung, Linver, & Brooks-Gunn, 2002). For instance, Garrett et al. (1994) analyzed the impact of poverty variables and family characteristics on the quality of the home environment (measured by the Home Observation for Measurement of the Environment [HOME]; Caldwell & Bradley, 1984). They found that being born into poverty and the duration and depth of poverty adversely affected the quality of children's home environments. Yeung et al. (2002) considered parental investments, such as the provision of cognitive stimulating materials, and warm and punitive parenting practices, as mediating associations between family income and the developmental outcomes of 3- to 5-year-old children. Although family income was directly associated with children's language and behavioral outcomes, the relationship was mostly mediated by maternal distress and parenting practices. Duncan et al. (1994) found that family income and poverty status predicted cognitive and behavioral outcomes in preschoolers, with effects being largely mediated by the quality of children's home environments.

Remaining gaps. Together, this new wave of studies offers insight into the pathways through which family economic resources and parenting quality jointly influence children's achievement. However, despite recent attempts to document these dual influences, such efforts remain rare and several noteworthy gaps persist.

First, with a few notable exceptions, studies that jointly consider the influences of parenting quality and economic resources on children's outcomes are based on limited measures of one or the other construct. Many studies in the economics literature define parenting in terms of expenditures on children or the amount of time spent with children. Most typically, these indicators of parenting derive from surveys and self-report rather than actual observations of the quality of parent–child interactions (as are more typical in developmental studies). Similarly, studies of family resources and child development often use measures that are not temporally linked to children's age. For example, measures of family income are often based on data from time frames that are either too narrow (e.g., 1 month) or overly broad (e.g., average family income across a span of years). Family income might fluctuate significantly over time due to changes in family structure, the work status of family members, or extraordinary circumstances, such as illness, and therefore should be measured during the specific age at which child development is being assessed. Furthermore, monetary resources are not the only resources that might influence children's cognitive development. Family structure, father residency, and parents' education and language skills have been found to contribute to children's development and academic achievement (Dahl & Lochner, 2005) and should be included in models that examine family resources and child development.

Second, few developmental studies offer a rigorous test of the contribution of parenting to children's outcomes. In particular, issues of selection bias or endogeneity are often not taken into account in analyses. For instance, parents with higher cognitive ability might be able to provide better parenting and also offer more resources to their children (in the form of income, the neighborhood where they choose to live, etc.). However, such characteristics of parents are not usually accounted for in analyses, thus creating selection bias in the estimates of the effects of parenting on children's outcomes. In addition, the associations between parenting practices and children's outcomes are reciprocal. With few exceptions, developmental studies do not regard both children's outcomes and parenting as potential dependent variables. In the absence of rigorous consideration of selection bias and endogeneity, developmental work is at risk of being ignored in the larger social science community (McCartney, Bub, & Burchinal, 2006).

Third, little is known about the contributions of family resources and parenting at *early developmental stages*. Family income has been found to most strongly influence children in the 1st years of life and has led to the recommendation that income-related policies target families with preschoolers (Duncan & Magnuson, 2005). Beyond family income, parent–child interactions shape the long-term trajectories of children already from infancy. For instance, children's literacy experiences influence their language and cognitive development by 14 months of age (Raikes et al., 2006; Rodriguez, Tamis-LeMonda, & Spellmann, 2005), and children's language and cognitive performance are stable from infancy to later ages (Adams, 1990; Campbell & Ramey, 1994; Magnuson, 2005; Mayer, 1997). However, most studies on family resources and parenting are based on families with school-aged children and adolescents. The few studies of "early childhood" include families with children older than 3 years or those with children across a span of ages (e.g., birth to 5 years). Such studies preclude precise statements about the ages at which parenting and family resources affect children in the first years of life.

Fourth, the effects of family resources and parenting on children's development have been typically examined at a single point in time, thereby limiting understanding of the dynamic processes that shape children's abilities across early development. An exception is longitudinal work showing that children who experienced low family income through early childhood showed worse academic achievement outcomes than children who experienced low levels of family income during middle childhood or adolescence (Duncan et al., 1998).

Longitudinal research also sheds light on reciprocal child–environment influences that might feed into children's developing abilities. Forty years ago, Bell (1968) argued against the dominant view in psychology that socialization was a parent-to-child process. He cited convincing research from human and animal studies that indicated the ways in which characteristics of offspring (ranging from physical appearance to skills and behaviors) evoked different responses in parents, which fed back into the development of their offspring. Ideas of reciprocity are further emphasized in current transactional models (e.g., Sameroff & MacKenzie, 2003), which posit that developmental outcomes are the product of continuous, dynamic interactions between children and their environments. In line with a transactional perspective, earlier measures of children's development should affect later measures of parenting just as earlier measures of parenting should affect later measures of children's development.

The limited focus on dynamic family–child influences is particularly evident in studies of low-income families. Studies based on nationally representative data implicitly assume that parenting quality in low-income families influences children's development and mediates the effects of family economic conditions in similar ways as in middle- and upper income families. However, this assumption remains to be tested. In addition, policy concerns around the cognitive and language delays of children in poverty (Brooks-Gunn & Duncan, 1997; Hoff, Laursen, & Tardif, 2002; Votruba-Drzal, 2003) arise soon in development, with disparities already evident by 3 years of age (Ackerman, Brown, & Izard, 2004; Hart & Risley, 1995; Hirsh-Pasek & Golinkoff, 2002; Hoff, 2003; NICHD Early Child Care Research Network, 2001). It is therefore important to take a dynamic approach to examining reciprocal processes in low-income families.

The Current Study

In the current study, we take an integrative approach to understanding the contributions of family income and parenting quality to children's cognitive development in the first 3 years. A main contribution is the rigorous testing of parenting effects on children's development (and vice versa) above the influences of family resources and children's earlier abilities. We consider that family resources affect children's development directly as well as indirectly through the quality of parenting behaviors. We address gaps of previous studies by (a) integrating comprehensive measures of parenting quality and family economic resources that are linked to children's ages in analyses of family influences on children's development, (b) testing for selection bias and considering influences of children's development outcomes on parenting quality, (c) focusing on children's developmental outcomes during the first 3 years, and (d) taking a dynamic, transactional approach to analyses by considering lagged, reciprocal influences between parenting quality and children's development outcomes.

To these ends, we examined a group of low-income families and their children who participated in the Early Head Start (EHS) Research and Evaluation Study. Families were assessed at four specific points in early development (baseline, 14, 24, and 36 months). Our analyses include measures of family economic resources that are developmentally linked to children's age and measures of observed parenting quality. We emphasize processes in low-income families and explore reciprocal effects of family resources, parenting quality, and child development outcomes at adjacent ages.

Method

Participants

Participants were 2,089 mothers and their children drawn from the EHS Research and Evaluation Study.[1] The EHS project was a randomized control experimental evaluation conducted in 17 sites across the United States. Study participants were low-income families who had sought assistance from local EHS community agencies between 1996 and 1999. Data on family characteristics and functioning, parenting and children's health, and cognitive outcomes were obtained when mothers applied to EHS (i.e., baseline), as well as from interviews, observations, and direct child assessments when children were 14, 24, 36 months. These interviews were augmented with telephone interviews at 6, 15, and 26 months after the baseline interview.

The EHS Research and Evaluation Study originally recruited 3,001 mothers and their children. However, 355 of the initially recruited participants dropped out of the study soon after being selected and therefore had virtually no data. Of the remaining 2,646 families for whom some data existed, we included 2,089 families who had data on children's outcomes, parenting quality, and family income from *at least one of the three assessments*. This sample did not differ from the sample of families who did not have at least one measure of parenting, one child outcome, and one measure of family income (N = 557), with the exception that the participants without any data on these measures were more likely to be Black.[2]

Table 1 presents demographic characteristic of the sample of 2,089 low-income children and their families. Almost 71% of the families in the sample had annual income below the poverty line at the time of recruitment, and the average annual family income reported was $11,532.4 (in 2003 dollars). About one fifth of the mothers were adolescents when their children were born and all mothers averaged 22.4 years. At baseline, about a quarter of the mothers were married and living with their husbands. Slightly more than half of the children were male and more than 60% were firstborns. The mothers in the sample were from White, African American, and Hispanic backgrounds. About half of the mothers had at least a General Educational Development Test (GED) or high school degree at baseline and subsequent assessments. Most of the mothers reported that English was their first language.

Procedures

Home visits at 14, 24, and 36 months consisted of a 45-min parent interview, direct assessments of

Table 1
Demographics

Variable	Percent	Frequency (*n*)	Missing in sample	Valid percent
In EHS program	51.5	1,075	0	51.5
Focus child is male	50.7	1,059	0	50.7
Focus child was firstborn	61.4	1,283	9	61.7
Birth weight < 2,500 g	7.4	155	263	8.5
Mother's ethnicity				
White	40.2	839	0	40.2
Black	34.8	728	0	34.8
Hispanic	25.0	522	0	25.0
Teen mothers	20.6	430	0	20.6
Mother's primary language is English	77.3	1,615	75	80.2
Mother's education at baseline				
< 12 years	44.3	926	78	46.0
= 12 years	28.7	599	78	29.8
> 12 years	23.3	486	78	24.2
Mother's education at 14 months				
< 12 years	38.9	813	189	42.8
= 12 years	31.5	657	189	34.6
> 12 years	20.6	430	189	22.6
Mother's education at 24 months				
< 12 years	33.5	699	302	39.1
= 12 years	31.6	660	302	36.9
> 12 years	20.5	428	302	24.0
Mother's education at 36 months				
< 12 years	30.5	637	351	36.7
= 12 years	31.8	664	351	38.2
> 12 years	20.9	437	351	25.1
Living arrangements at baseline				
Mother is married and living with husband	25.7	536	3	25.7
Mother lives with other adults	39.0	815	3	39.1
Mother lives alone	35.2	735	3	35.2
Region of residence				
Northeast	21.4	448	0	21.4
Midwest	25.9	542	0	25.9
South	17.3	362	0	17.3
West	35.3	737	0	35.3
Families with annual income below poverty line	70.9	2,127	534	86.2

Note. EHS = Early Head Start.

children's cognitive abilities, a 10-min videotaped session of semistructured mother–child free play, and a videotaped dyadic challenge session (highchair task at 14 months, teaching tasks at 24 and 36 months). Interviewers also completed a checklist on their observations of the home environment at the end of each visit. At the 24-month assessment, mothers' cognitive and language skills were assessed using the Woodcock–Johnson (WJ) Picture Vocabulary Test (WJ–III Tests of Achievement [Standard Battery]; Woodcock & Johnson, 1989).

For the *free play sessions*, mothers were presented with three bags containing a book and a standard set of age-appropriate toys. At the 14- and 24-month assessments, the bags contained a book (*Good Dog Carl* and *The Very Hungry Caterpillar*, respectively), a cooking set, and a Noah's Ark set with animals. At the 36-month assessment, the bags contained a book (*The Very Hungry Caterpillar*), a cash register and grocery items, and a container of interlocking blocks. Mothers were instructed to begin with Bag 1 and to finish with Bag 3 and were encouraged to play with their children as they ordinarily would and to face the camera.

The challenge task at 14 months consisted of mother placing her infant in a booster chair, with the infant's back toward the mother. Mothers sat behind the infant and were asked to draw a picture of the family while the infant remained in the chair without any objects for a period of 4 min. Mothers were told "I'd like you to put [Child's Name] in this chair and strap him/her in. You will then sit in this chair behind your baby. We'd like your child to remain in the chair for 4 minutes while you draw a picture of your family. Feel free to take care of your child as you normally would, but we ask that you not provide him/her with any objects and that you not remove him/her from the chair." The challenge task at 24 months consisted of mothers teaching their toddlers how to sort blocks according to color or learn words for articles of clothing shown in a book. Mothers selected the task that they felt their child did not know. Once they selected the task, they were asked to sit on a mat and to let the experimenter know when they had finished teaching the child. The challenge task at 36 months consisted of the child being presented with three puzzles of increasing difficulty one at a time. Mothers were told that the puzzles were challenging for 3-year-olds. They were instructed to first let the child work on each puzzle alone, and then, they could give the child the help they felt the child needed. All mother and child interactions were videotaped and subsequently coded by staff at the Center for Children and Families, Columbia University, Teachers College.

Measures

All measures in this study, except family income, were based on data obtained during home visits. Family income was measured at baseline and at the telephone interviews conducted at 6, 15, and 26 months after the baseline interview. These income data were reconstructed to create an income timeline that enabled us to estimate household annual income for the periods leading up to each of the three child ages (14, 24, and 36 months).

Child cognitive outcomes. At the 14-, 24-, and 36-month assessments, cognitive development was assessed with the Mental Development Index (MDI) of the Bayley Scale for Infant Development, Second Edition (BSID – II; Bayley, 1993). The MDI assesses vocalizations, language skills, memory, problem solving, early number concepts, classification abilities, generalization, and social skills. This indexed scale has a mean of 100 and a standard deviation of 15. The MDI has shown adequate internal consistency and interrater reliability.[3] A score of 85 – 100 in this scale is within normal limits.

Background factors. Several background factors were included as controls for initial conditions of mother and child. These were conceptualized as characteristics of mothers and children that do not change over time and are likely to influence parenting quality and children's outcomes. Background factors were children's gender, birth weight, mother's age at childbirth, ethnicity, primary language, and mothers' cognitive skills as indexed by the WJ.

Children's characteristics such as gender and birth weight might play an important role in determining parenting quality and children's developmental outcomes. For example, it has been found that girls present a slight advantage over boys during early stages of cognitive and language development (Bornstein, 2002; Fenson et al., 1994). In addition, it has been found that birth weight significantly predicts later development (Bradley, Whiteside, Mundfrom, & Blevins-Knabe, 1995; Landry et al., 2006; Shenkin, Starr, & Deary, 2004).

Maternal characteristics and skills influence the quality of parenting styles and child development outcomes (Bornstein, 2002). For instance, teenage mothers tend to be less mature and lack a well-established maternal self-definition (Brooks-Gunn & Chase-Lansdale, 2002) and their children perform less well than those of older mothers on measures of cognitive and language development (Keown, Woodward, & Field, 2001; Rodriguez et al., 2005). Mothers' ethnicity and/or primary language relates to how responsive or intrusive mothers are in their interactions with the children (Ispa et al., 2004) and to maternal teaching styles and shared narratives in the preschool years (Melzi, Paratore, & Krol-Sinclair, 2000; Paratore, Melzi, & Krol-Sinclair, 2003). Finally, measures of mothers' cognitive abilities, as assessed through standardized instruments, relate to parenting and children's cognitive outcomes (e.g., Yeung et al., 2002).

Here, dichotomous indicators were used for child gender, birth weight (< 2,500 g vs. other), mothers' age (< 18 years old at child's birth vs. other), ethnicity (contrasts for African American vs. other and for Hispanic vs. other), and mothers' primary language (1 indicating English).

Family resources. We considered four measures of family resources: mother's own reading frequency, mother's educational achievement, parental living arrangements, and income per capita in the family. Summary statistics of these measures are presented in Table 3. Although these four measures do not fully capture a family's human capital (conceptualized as all resources that enable individuals to achieve well-being), they each contribute in significant ways to children's and family's well-being. For instance, mother's own reading frequency and educational achievement indicate maternal ability to teach and stimulate learning. Father's residency reflects the availability of additional financial and parenting resources.

Income per capita in the family relates to wealth and consumption resources allocated to children.

Mothers' educational achievement was represented by mothers' years of education at the time of the assessment. Mother's own reading frequency was assessed with a single item coded on a 5-point Likert scale (5 = *daily reading*, 1 = *reading a few times per year*, 0 = *never reads*). Father residency was coded dichotomously at each assessment, representing whether the biological father lived with the mother and child at the 14-, 24-, and 36-month assessments.

Family income was measured at baseline or recruitment and at the assessments that took place at 6, 15, and 26 months after the baseline interview. Given the goal of including repeated measures of annual monetary resources that temporally aligned with the child assessments, we computed the amount of money received by all family members (including the mother) for the periods in which each child in the sample was 0 – 14 months, 14 – 24 months old, and 24 – 36 months. Therefore, family income represented the total amount of money received by the mother and the members of the family, as reported by the mother. It included public assistance and other transfers (e.g., food stamps) because the question was not limited to the amount of money *earned* by family members but excluded taxes and deductions. From these multiple assessments, monthly family income was calculated for each of the 36 months from the child's birth date through the child's third birthday. These month-by-month calculations were summed to create totals for 0 – 14 months, 14 – 24 months, and 24 – 36 months income figures. Whenever data were not available for a portion of a development period, the available monthly income data were prorated so as to span the relevant developmental periods. To provide a more accurate portrayal of the amount of monetary resources available to children at each age, income per capita was used in analyses.

Parenting quality. Parenting quality was based on a composite score of several dimensions of parenting that were assessed from three protocols at each age: videotaped mother – child play (three-bag task), videotaped challenge tasks (high chair and teaching), and the HOME scale. First, maternal supportiveness was assessed from free play at 14, 24, and 36 months. Supportiveness was the average of mothers' ratings on three Likert-scaled items: maternal sensitivity, positive regard, and cognitive stimulation (1 = *very low* to 7 = *very high* on each dimension), following the established precedent of the EHS Consortium studies (e.g., Berlin, Brady-Smith, & Brooks-Gunn, 2002; Love et al., 2005; Shears & Robinson, 2005; Tamis-LeMonda et al., 2004). Sensitivity reflected the extent to which the mother was responsive to her child's distress and nondistress cues; positive regard reflected mothers' demonstration of love, respect, and admiration toward her child; and cognitive stimulation reflected mothers' attempts to teach her child so as to enhance perceptual, cognitive, and linguistic development. These items covaried strongly at all ages (r ranges from .60 to .62, .53 to .68, and .50 to .71 at 14, 24, and 36 months, respectively. All were significant at the .01 level). The supportiveness composite also showed adequate internal consistency at each age (Cronbach's αs = .82, .83, and .82 at 14, 24, and 36 months, respectively).

Second, interviewers completed the HOME scale (Caldwell & Bradley, 1984), and total HOME scores were used. This scale rates the quality of the home environment including provision of learning materials, activities (e.g., reading), and observations of mother and child verbal interactions during the full duration of the home visit.

Finally, mother – child interactions were coded from the videotaped challenge task at each age (high-chair task at 14 months, teaching task at 24 months, and puzzle task at 36 months). At 14 months, maternal sensitivity and positive regard were coded on 7-point Likert scales and summed to create a composite supportiveness score. At 24 months, mothers' teaching was coded using the teaching scale of the Nursing Child Assessment Satellite Training (NCAST; Barnard, 1976). This scale coded a total of 50 binary items reflecting mother's sensitivity, responsiveness to child distress cues, cognitive growth fostering, and social-emotional growth fostering. These items were assessed as present or absent and summed to a total score (Barnard, 1976). At 36 months, scores on mothers' supportive presence and quality of maternal assistance were rated during the teaching and puzzle tasks (1 = *very low* to 7 = *very high*). Supportive presence measured the mother's emotional availability and physical and affective presence during the task and quality of assistance focuses on the instrumental support and assistance provided by the mother during the task.

Because the coding of the three parenting protocols yielded scores on different metrics, scores were z transformed and then summed into a total parenting quality score at each age. Bivariate correlations between the three parenting quality scores were significant at the $p < .01$ level at all ages (r ranges from .31 to .45, .30 to .42, and .41 to .51 at 14, 24, and 36 months, respectively). Confirmatory factor analysis was then run to verify that the three measures of parenting loaded onto a single factor at each of the three ages. Table 2 presents the results from this analysis. The strongest loadings were obtained for

Table 2
Factor Loadings of Confirmatory Factor Analysis of Parenting Quality Aspects

Measure and variable	Unstandardized factor loading	Standardized factor loading	R^2
14-month parenting quality			
14-month Maternal Supportiveness score	1.00[a]	0.70	.49
14-month HOME score	3.03*	0.63	.40
14-month High Chair Task score	0.99*	0.48	.23
24-month parenting quality			
24-month Maternal Supportiveness score	1.00[a]	0.69	.48
24-month HOME score	2.94*	0.62	.37
24-month Teaching Task score	2.82*	0.50	.25
36-month parenting quality			
36-month Maternal Supportiveness score	1.00[a]	0.71	.50
36-month HOME score	4.19*	0.58	.34
36-month Puzzle (Teaching) Task score	1.21*	0.72	.52

Note. HOME = Home Observation for Measurement of the Environment.
[a]Not tested.
$^*p \leq .05$.

the HOME score at all ages. However, as indicated by the standardized estimates, the largest correlations with the parenting factor were seen for maternal supportiveness during free play at 14 and 24 months and by supportiveness in the teaching task at 36 months. All the unstandardized estimates that were not normalized to be 1 were significant at a level of $p <$ .05. As indicated by the R^2 of each indicator, parenting quality explained between 23.4% and 51.7% of the variances of the indicators. Because all factor loadings were significant, the creation of composite constructs of parenting at each age was deemed to be appropriate. The parenting quality composite at each age showed sufficient internal consistency (Cronbach's αs = .63, .62, and .70 at 14, 24, and 36 months, respectively).

Results

We first present descriptive statistics of background measures, parenting quality, family resources, and child cognitive performance, followed by bivariate associations among these measures. Next, structural equation modeling (SEM) tests associations among measures across the three ages. The conceptual model that guided the structural equation analyses is presented in Figure 1. At each age (14, 24, and 36 months), we assumed that family resources (family income, father's residency, etc.) influence child outcomes both directly and indirectly (through parenting quality). Parenting quality was assumed to directly affect children's outcomes at each age. We controlled for earlier experiences by linking parenting quality at 24 and 36 months to previous parenting quality and by linking child cognitive performance at 24 and 36 months to previous child performance. We hypothesized reciprocal effects between children and parents. Thus, the model includes lagged effects from children's performance at each age to parenting quality at the adjacent age and vice versa (diagonal paths). Family background factors (e.g., mothers' and children's demographic characteristics) served as controls.[4]

Descriptive Statistics

Descriptive statistics for all measures and constructs appear in Tables 3 and 4. As indicated in Table 3, mothers averaged 89 on the WJ Test of Achievement, which is approximately 0.75 *SD* below the population norm. Children's Bayley MDI scores were within the normal limits but below population norms at all ages. Bayley scores declined significantly from 14 to 24 and 36 months, as indicated in a repeated measures analysis of variance. This drop may reflect changes to the skills being tapped by the Bayley MDI at earlier versus later ages. That is, the Bayley MDI at 14 months in part measures motor skills, whereas at 24 and 36 months, Bayley items increasingly tap language development.

Table 4 presents summary statistics of mothers' own reading frequency, family income, and parenting quality. On average, mothers were reading on their own between a few times a month and once a week. Sample averages of total family income and income per capita (all in 2003 dollars) for the considered developmental periods were stable over time. The averages at 24 months shown in Table 4 appear to be lower than at other ages because family income amounts at 24 months represented economic resources for the period between 14 and 24 months, which is only a 10-month period. The average income per capita ranges from about $4,600 during the 0- to 14-month period to about $5,100 in the 24- to 36-month period, with average family size between four and five family members. Although summary statistics are presented for both total and per capita income, analyses were based on income per capita.

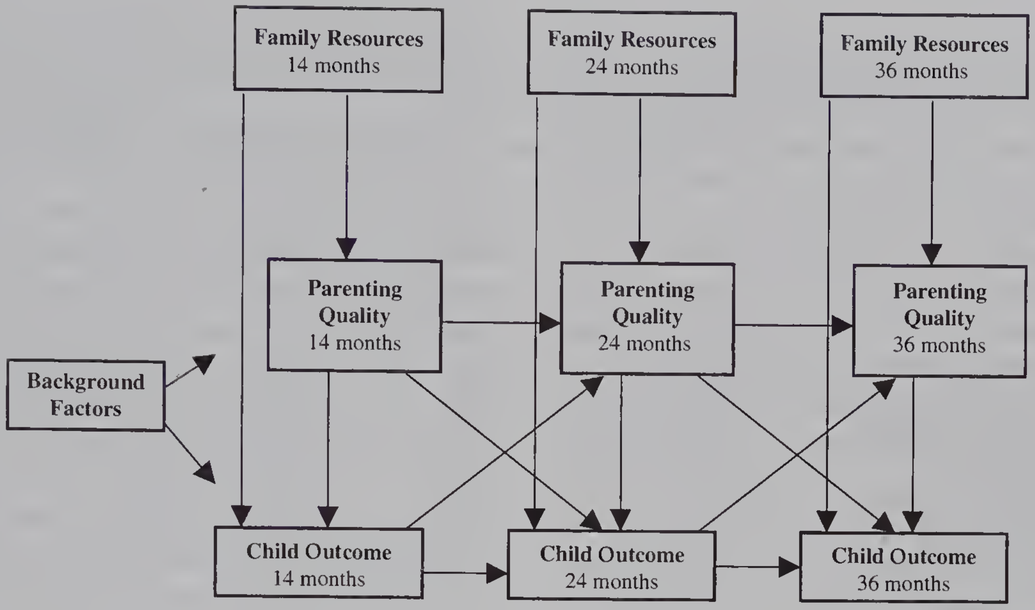

Family Resources 14 months
Family Resources 24 months
Family Resources 36 months
Parenting Quality 14 months
Parenting Quality 24 months
Parenting Quality 36 months
Background Factors
Child Outcome 14 months
Child Outcome 24 months
Child Outcome 36 months

Figure 1. Conceptual model.

Bivariate Associations

Table 5 presents bivariate correlations between the background factors and the parenting quality score at each age. The strongest associations are between mothers' scores on the WJ Test of Achievement and parenting quality at each age. The maternal primary language indicator (=1 if the primary language is English) also relates to parenting quality, but associations decrease over time (*r*s = .20, .13, and .11, at 14, 24, and 36 months, respectively, all with $p \leq .01$). The indicators for being a teenager, African American, and Hispanic are all negatively and significantly associated with the quality of parenting score but again decrease over time. Neither child gender nor low birth weight related to parenting quality.

We calculated bivariate correlations between income per capita in the family, mother's years of education, mother's own reading frequency, father residency, and parenting quality at each age. Income per capita was positively associated with all variables at each age. Mothers' own reading frequency was very modestly associated with income per capita in the family at each age ($r = .06$, $p \leq .05$ at 14 months; $r = .11$, $p \leq .01$ at 24 months; $r = .07$, $p \leq .01$ at 36 months). Income per capita related most strongly to parenting quality (*r*s = .26, .25, and .28, at 14, 24, and 36 months, respectively, all with $p \leq .01$). The parenting quality score was also associated with all other variables at each age, with associations being strongest to mothers' years of education (*r*s = .45, .37, and .38, at 14, 24, and 36 months, respectively, all with $p \leq .01$).

Table 6 presents the associations between background factors, family resources, and parenting quality scores and children's Bayley MDI scores at each age. Both family resources and parenting scores correlated with children's Bayley scores at all ages. The weakest associations were between mothers' background factors at 14 months and children's Bayley scores at 14 months and between children's

Table 3
Background Factors

Variable	Valid *N*	*M*	*SD*	Minimum	Maximum
Male focus child indicator	2,089	0.51	0.50	0	1
Low birth weight indicator	1,826	0.08	0.28	0	1
Teen mother indicator	2,089	0.21	0.40	0	1
African American indicator	2,089	0.35	0.48	0	1
Hispanic indicator	2,089	0.25	0.43	0	1
Mother's primary language is English indicator	2,014	0.80	0.40	0	1
Mother's score on WJ Test	1,477	89.15	11.62	50.00	130.00

Note. WJ = Woodcock–Johnson.

Table 4
Child Outcomes, Family Resources, and Parenting Quality

Variable	Valid *N*	*M*	*SD*	Minimum	Maximum
Child outcomes					
14 months Bayley MDI	1,642	98.26	11.21	49	130
24 months Bayley MDI	1,585	89.22	13.61	49	134
36 months Bayley MDI	1,485	90.79	12.57	49	134
Mother's years of education					
Completed at 14-month assessment	1,280	11.24	2.71	0	19
Completed at 24-month assessment	1,614	11.44	2.61	0	20
Completed at 36-month assessment	1,712	11.50	2.60	1	20
Mother's own reading frequency					
14 months reading frequency score	1,907	3.29	1.73	0	5
24 months reading frequency score	1,798	3.36	1.71	0	5
36 months reading frequency score	1,752	3.46	1.65	0	5
Family income (tens of thousands of dollars)					
0–14 months income	1,507	1.84	1.40	0.03	13.70
14–24 months income	1,910	1.46	1.12	0.01	9.41
24–36 months income	1,775	1.98	1.45	0.01	11.04
Family income (tens of thousands of dollars)					
0–14 months income per capita	1,415	0.46	0.37	0.01	4.39
14–24 months income per capita	1,702	0.37	0.29	0.00	2.37
24–36 months income per capita	1,632	0.51	0.42	0.00	4.92
Father lives with mother and child					
14 months resident father	1,905	0.48	0.50	0	1
24 months resident father	1,750	0.47	0.50	0	1
36 months resident father	1,673	0.46	0.50	0	1
Parenting quality score[a]					
14 months parenting quality score	1,543	—	2.29	−7.48	6.67
24 months parenting quality score	1,494	—	2.26	−9.65	5.59
36 months parenting quality score	1,442	—	2.35	−7.49	6.19

Note. MDI = Mental Development Index.
[a] Sum of *z*-score values with a mean of 0.

Bayley scores at 14 months and later family resources at 24 and 36 months.

Finally, income per capita, parenting quality, and children's cognitive development outcomes were stable across the three assessments, as suggested by bivariate correlations of each variable at a given time with itself at the subsequent assessment. In particular, parenting quality was strongly stable and stability

Table 5
Bivariate Correlations of Background Variables to Parenting Quality

	Parenting quality		
Variable	14 months	24 months	36 months
Male child indicator	−.06*	−.05	−.05
Low birth weight indicator	−.03	−.05	−.04
Teen mother indicator	−.16**	−.16**	−.16**
African American indicator	−.24**	−.23**	−.22**
Hispanic indicator	−.20**	−.13**	−.11**
Mother's primary language is English indicator	.20**	.13**	.11**
Mother's score on WJ Test	.54**	.46**	.48**

Note. WJ = Woodcock–Johnson.
$^{*}p \leq .05$. $^{**}p \leq .01$.

coefficients increase over time ($rs = .62, .63$, and .65, for comparisons of 14–24, 14–36, and 24–36 months, respectively, all with $p \leq .01$).

Together, these findings indicate that background factors and family resources are associated with parenting quality at each age and that background factors, family resources, and parenting quality consistently predict children's Bayley scores. Thus, testing a model that includes repeated assessments of these sets of measures is justified. Moreover, in light of the high stability in child cognitive performance and parenting quality across 14-, 24-, and 36-month assessments, the inclusion of earlier measures of child performance and parenting quality produces a very conservative estimate of the effects of parenting and income on children's performance at later ages.

Estimation Strategy

SEM was used to test paths between background factors, family resources, parenting quality, and child outcomes across the three assessments. This estimation approach simultaneously tests associations between all measures in the proposed model (Schumacker & Lomax, 1996; Yeung et al., 2002). In addition, SEM estimates direct and indirect effects of predictors in a longitudinal design. We used the Mplus (Muthén & Muthén, 2004) program to estimate our model. Mplus uses maximum likelihood and permits estimation of structural equation models for which some or all variables have missing data. In this estimation procedure, the likelihood function is conditioned to the missing data in the model, and the nonmissing data are used to find the parameters that maximize the conditioned likelihood function. Thus, this estimation method uses all the information that is available from the data. This technique has been shown to yield better estimates than estimation methods that rely on case deletion or single imputation (McCartney et al., 2006; Schafer & Graham, 2002).

Model Fit

We present several statistical evaluations of the model in Table 7. Specification 1 reports results of a model in which direct paths from the family resources variables (i.e., family income per capita, mother's years of education, father's residency, and

Table 6

Bivariate Correlations of All Variables With Bayley MDI 14, 24, and 36 Months

	Bayley MDI		
Variable	14 months	24 months	36 months
Male child indicator	−.09**	−.11**	−.08**
Low birth weight indicator	−.15**	−.05*	−.09**
Teen mother indicator	−.01	−.07**	−.05*
African American indicator	−.03	−.11**	−.18**
Hispanic indicator	−.01	−.11**	−.07**
Mother's primary language is English indicator	−.01	−.07**	−.05**
Mother's score on WJ Test	.16**	.38**	.32**
Income per capita at 14 months	.09**	.17**	.16**
Mother's years of education at 14 months	.12**	.24**	.21**
Mother's own reading frequency at 14 months	.07**	.12**	.09**
Father lives with mother and child at 14 months	.04	.06*	.07*
Parenting quality score at 14 months	.26**	.40**	.39**
Income per capita at 24 months	.06*	.18**	.15**
Mother's years of education at 24 months	.11**	.25**	.22**
Mother's own reading frequency at 24 months	.04	.10**	.03
Father lives with mother and child at 24 months	.04	.07**	.04
Parenting quality score at 24 months	.21**	.38**	.40**
Income per capita at 36 months	.09**	.10**	.14**
Mother's years of education at 36 months	.09**	.24**	.21**
Mother's own reading frequency at 36 months	.04	.09**	.09**
Father lives with mother and child at 36 months	.02	.08**	.07**
Parenting quality score at 36 months	.18**	.36**	.41**

MDI = Mental Development Index; WJ = Woodcock–Johnson.
$^*p \leq .05$. $^{**}p \leq .01$.

mothers' reading frequency) to the children's outcomes at each age are *excluded*. Specification 2 reports the same results for a specification in which all paths from the model presented in Figure 1 are included. Disturbance terms and unanalyzed associations in background factors and resources variables are not depicted but were included in all estimations.

Because Specifications 1 and 2 represent nested models, we performed a likelihood ratio test to assess which model provided a better fit. The test statistic shown in Table 7 (a χ^2 value with 12 *df*) indicates that at a .05 significance level, it is not possible to reject the null hypothesis that the paths of the family resources variables to the child cognitive outcomes are statistically equal to zero (based on a critical χ^2 value of 21.03). This result suggests that the effects of family resources on child outcomes are completely mediated by parenting quality at each age.

We also tested Specification 1 against a model that excluded the paths between past child outcomes and current parenting and between past parenting quality and current child outcomes (Specification 3). Similarly to Specification 1, Specification 3 excludes direct paths from family resources variables to child outcomes. Again, a likelihood ratio allowed us to assess whether the reciprocal paths between child outcomes and parenting quality significantly contribute to a better model fit. The test statistic (a χ^2 value with 4 *df*) indicates that at a .05 significance level, we can reject the null hypothesis that the paths of past child outcomes to current parenting quality and paths from past parenting to present child outcomes are statistically equal to zero (based on a critical χ^2 value of 9.49). This result suggests that reciprocal effects between children and parents exist.

Given the results of these tests, we next adopted the more parsimonious specification of the model that excludes direct paths from the family resources variables to the child outcomes at each age but considers reciprocal paths between child outcomes and parenting quality (Specification 1).

Overall, the fit indices indicate that Specification 1 provides a better fit than Specifications 2 and 3. The standardized root mean square residual and the root mean square error of approximation for Specification 1 are below .05, which can be considered a good fit. The Tucker–Lewis index (TLI) and the comparative fit index are above .90 for Specification 1, which in general is considered a good fit. The TLI for Specifications 2 and 3 are slightly below .90. The chi-square values of model fit were also computed for the three specifications. However, because this statistic is sensitive to the sample size (N = 2,089 in the present study), its value for all specifications was significant. Given that all other fit indices reported in the present study take into account the value of the chi-square and the degrees of freedom in our model, we considered that the values obtained from these indices provide enough evidence that the model we propose fits the data well.

Finally, the proportion of explained variance (R^2) for each of the dependent variables in the model also provides a measure of model fit. The proportion of explained variance of the Bayley MDI at 14 months is close to 10% and it increases to 32% and 40% at 24 and 36 months, respectively. The model also explains a significant proportion of the variance of parenting quality at each age. In all specifications, the proportion of explained variance across parenting quality scores ranges from 38% at 14 months to almost 50% at 36 months. Increasing proportions of explained variance across time suggest cumulative

Table 7
Model Fit

Specification	SRMR	RMSEA	TLI	CFI	Log likelihood	Chi-square value of log-likelihood difference
1. Direct paths from family resources to child outcomes are excluded	.02	.04	.90	.93	−57,303.43	
2. Direct paths from family resources to child outcomes are included	.02	.04	.89	.93	−57,296.49	
Difference between Specifications 2 and 1					6.94	13.89
3. Reciprocal paths from child outcomes to parenting and vice versa are excluded	.03	.04	.89	.91	−57,335.26	
Difference between Specifications 3 and 1					31.83	63.64

Note. The differences in log likelihoods are absolute values. SRMR = standardized root mean square residual; RMSEA = root mean square error of approximation; TLI = Tucker–Lewis index; CFI = comparative fit index.

influences in children's cognitive development and parenting quality.

Parameter Estimates

Figure 2 presents standardized (in bold font) and unstandardized estimates of the paths of the model in which the effects of family resources variables on child outcomes are mediated by the quality of parenting (Specification 1).[5]

At all ages, the largest effect on parenting quality from a family resource variable was obtained for mothers' years of education, with higher maternal education being associated with higher parenting quality. Fathers' residency and mothers' reading frequency were positively associated with parenting quality at 24 and 36 months (net of other measures in the model), although effect sizes were small. Higher income per capita in the family is also related to higher parenting quality, although the associations were small relative to others in the model. The effects of the family resources variables on parenting quality were small at all ages. However, these effects remained significant at 24 and 36 months even after controlling for past parenting quality. The strongest effects in the model were the associations between current and past parenting quality, revealing high levels of stability being already established at these early child ages. The standardized estimate of the path from parenting quality at 14 months to parenting quality at 24 months was .52, and the estimate of the path from 24 months to 36 months parenting quality was .53. Parenting quality at each age had a positive direct effect on child cognitive outcomes. The effects of parenting quality on child outcomes were moderate, but these effects were still significant even when the strongest effects on child cognitive outcomes were included—namely, prior child outcomes. Finally, lagged parenting quality was positively associated with child outcomes at 24 and 36 months, and lagged child outcomes had positive effects on parenting quality at 24 and 36 months. Although these estimates were small, they suggest the presence of reciprocal effects between children's abilities and subsequent parenting behaviors under extremely conservative tests.

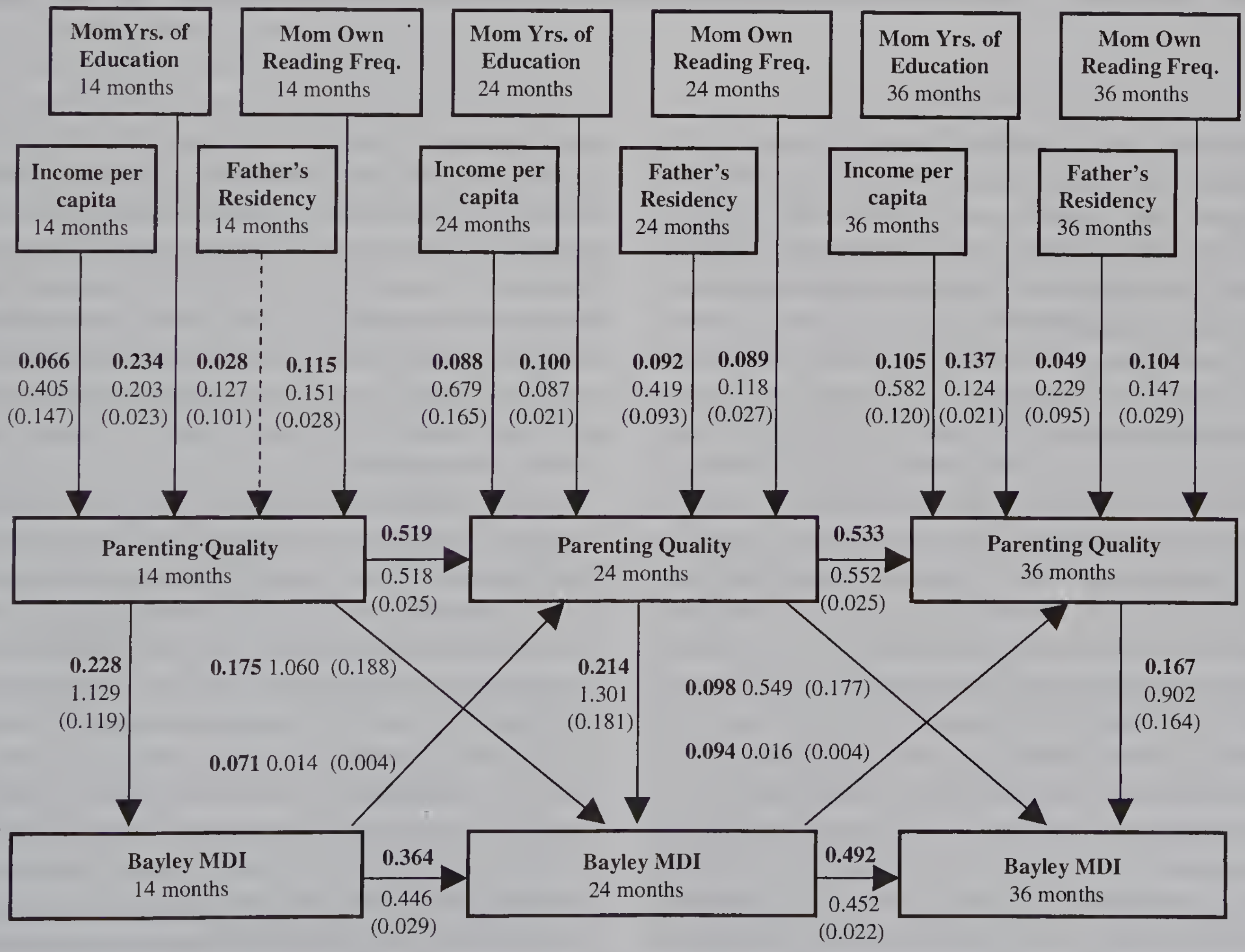

MomYrs. of Education 14 months
Mom Own Reading Freq. 14 months
Mom Yrs. of Education 24 months
Mom Own Reading Freq. 24 months
Mom Yrs. of Education 36 months
Mom Own Reading Freq. 36 months
Income per capita 14 months
Father's Residency 14 months
Income per capita 24 months
Father's Residency 24 months
Income per capita 36 months
Father's Residency 36 months
0.066 0.405 (0.147)
0.234 0.203 (0.023)
0.028 0.127 (0.101)
0.115 0.151 (0.028)
0.088 0.679 (0.165)
0.100 0.087 (0.021)
0.092 0.419 (0.093)
0.089 0.118 (0.027)
0.105 0.582 (0.120)
0.137 0.124 (0.021)
0.049 0.229 (0.095)
0.104 0.147 (0.029)
Parenting Quality 14 months
0.519 0.518 (0.025)
Parenting Quality 24 months
0.533 0.552 (0.025)
Parenting Quality 36 months
0.228 1.129 (0.119)
0.175 1.060 (0.188)
0.071 0.014 (0.004)
0.214 1.301 (0.181)
0.098 0.549 (0.177)
0.094 0.016 (0.004)
0.167 0.902 (0.164)
Bayley MDI 14 months
0.364 0.446 (0.029)
Bayley MDI 24 months
0.492 0.452 (0.022)
Bayley MDI 36 months

Figure 2. Standardized (bold) and unstandardized coefficient estimates (values given in parentheses are standard errors).
Note. Paths with solid lines are significant at the $p < .05$ level. This model excludes direct paths from family resources variables to child outcomes. MDI = Mental Development Index.

Table 8

Unstandardized Estimates for Indirect Effects of Family Resources and Background Variables on Child Outcomes

	Bayley MDI		
Variable	14 months	24 months	36 months
Income per capita at 14 months	.46*	—	—
Mother's years of education at 14 months	.23*	—	—
Mother's own reading frequency at 14 months	.17*	—	—
Father lives with mother and child at 14 months	.14	—	—
Income per capita at 24 months	—	.88*	—
Mother's years of education at 24 months	—	.11*	—
Mother's own reading frequency at 24 months	—	.15*	—
Father lives with mother and child at 24 months	—	.55*	—
Income per capita at 36 months	—	—	.53*
Mother's years of education at 36 months	—	—	.11*
Mother's own reading frequency at 36 months	—	—	.13*
Father lives with mother and child at 36 months	—	—	.21*
Male child indicator	−2.10*	−.97*	−.49*
Low birth weight indicator	−6.63*	−3.08*	−1.54*
Teen mother indicator	−.28*	−.57*	−.40*
African American indicator	−.16*	−2.32*	−1.66*
Hispanic indicator	−.35	−.71	−.50
Mother's primary language is English indicator	−.18	−.36	−.26
Mother's score on WJ Test	.07*	.14*	.10*

Note. Statistical significance of indirect effects is based on bootstrap standard errors. MDI = Mental Development Index; WJ = Woodcock–Johnson.
$^*p \leq .05$.

Table 8 presents estimates of the indirect effects of family resources variables on children's outcomes and of the total effects (i.e., direct and indirect [through parenting] effects) of background factors on children's outcomes. We report unstandardized estimates for interpretation purposes.[6] These unstandardized coefficients represent the effect of a unit increase in family resources variables and background factors on the child's Bayley MDI score. The estimates reported in Table 8 show small effects of a unit increase in family resources variables on child outcomes. For instance, an increase of 1 year in maternal education at 14 months produces an increase in the score on the Bayley MDI of less than one fourth of a point. Likewise, increases of 1 year in maternal education at 24 and 36 months create increases in the Bayley MDI score of about one ninth of a point. Similar effects on child outcomes were found in the case of mother's own reading habits and father's residency. The effect of income per capita on child cognitive outcomes was small and weakly increasing over time. According to the estimates presented in Table 8, an increase of $10,000 in the income per capita for the period between birth and 14 months of age would increase the score on the Bayley MDI by almost half a point. For the period between 14 and 24 months, a similar increase in income per capita would increase the score on the Bayley MDI at 24 months by close to a point. Finally, an increase of $10,000 in the income per capita for the period between 24 and 36 months of age would increase the score on the Bayley MDI at 36 months by about half of a point. Although each of these influences is small by itself, they are significant, and the total accumulation of these predictors on children's Bayley scores is meaningful.

We consider important to compare the effects of the family resources variables on child outcomes to the effects of the background factors. According to the results reported in Table 8, being a male child decreases the score on the Bayley MDI at 14 months by 2.1 points. At 24 and 36 months, the negative effects of child gender on child outcomes were smaller but still significant. Low birth weight had significant negative effects on cognitive outcomes at all ages. At 14 months, having low birth weight decreases the score on the Bayley MDI by 6.6 points. Although this negative effect declines over time, it was still significant at 36 months. Being a teen mother at the birth of the child and being African American had negative effects on the Bayley MDI scores at all ages. Being Hispanic and the mothers' primary language did not have significant effects on the child outcomes. Finally, mothers' language and cognitive skills, measured by their score on the WJ Picture Vocabulary Test, had small effects on the child cognitive outcomes at all ages. For example, at 14 months, an increase of 10 points in the mothers' WJ Test would increase the score on the Bayley MDI by close to a point. At 24 and 36 months, a similar increase in the mothers' WJ would create an increase in the Bayley MDI scores of 0.14 and 0.10, respectively. Regarding the effect of background variables on parenting quality, mothers with higher language and cognitive skills provided higher parenting quality. However, this effect was close to zero. Teenage and African American mothers tended to provide lower parenting quality.

Selection Bias

We were concerned about potential selection bias (from included and excluded variables) in our estimates. Economic or social resources might appear to affect child development outcomes but do so via their association to other measures not included in the model as well as unobserved characteristics and abilities of parents and children. Dahl and Lochner (2005) addressed this issue by using an estimation strategy that accounts for measurement error and omitted variable bias when analyzing the effects of income on children's math and reading outcomes. However, the authors did not consider the role of parenting quality and family resources other than income on determining child cognitive outcomes. NICHD Early Child Care Research Network and Duncan (2003) examined the effects of child care quality on children's cognitive development at preschool entry. As one way to assess whether selection bias might account for the identified associations between child care and child outcomes (e.g., characteristics in parents and children might underlie both child care choices and child measures), the authors ran a set of regression equations in which they progressively added controls to their models and compared estimates across models for their key associations. This analytic approach is based on earlier versions of the work of Altonji, Elder, and Taber (2005), which rests on the assumption that control variables included in nested models represent a subset of all possible controls. If the magnitude of the change in the coefficients of interest (e.g., child care to child outcome effects) is large, then selection bias is likely. If changes to coefficients are small or negligible, it suggests little additional bias from unobservable variables (Altonji et al., 2005).

Adopting this approach, we estimated changes to model coefficients for parenting effects in models that included the background variables in Figure 1 against models that included a larger set of controls. The additional controls were time invariant based on measures of parent and child characteristics gathered at the baseline assessment for the EHS Research and Evaluation Project as well as maternal depression measured by the Center for Epidemiological Studies Depression Scale (Radloff, 1977). These added controls included childbirth order, number of moves in the past year, family income as percentage of the poverty line, urban versus rural indicator, maternal employment status, maternal marital status, welfare receipt, and maternal education at baseline. The magnitude of the change of the effects of parenting quality on child outcomes in the contrasting models was negligible, even when measured in terms of standard deviations. (Maternal depression related inversely to child outcomes and the urban indicator related positively to child outcomes.) Parenting at all three assessments retained its significance in the fuller model. This suggests that there was little additional bias from unobserved variables.

Discussion

Despite widespread consensus that both parenting and economic resources matter for children's development, few studies have investigated the joint and unique influences of economic resources and parenting quality to children's development. This gap is particularly notable for infants and toddlers from low-income families. Furthermore, the dynamic nature of parenting behaviors and other family resources (e.g., mothers' education, father residency) is rarely examined. Thus, little is known about the effects of earlier experiences on current and later child outcomes and how child cognitive development reciprocally affects parenting behaviors over time.

In this study, we investigated the effects of parenting quality and family economic resources on the cognitive development of 0- to 3-year-old children from low-income families who participated in the EHS Research and Evaluation Study. The measure of parenting quality was based on direct observations of mother–child interactions and the home environment. With respect to family economic resources, we considered income per capita in the family during three developmental periods in the first 3 years, thereby matching family income to the time frame of child testing. In addition, models included other measures of families' resources such as maternal education, maternal reading habits, and fathers' residency. Finally, we analyzed mutual and dynamic influences of parenting quality and child cognitive outcomes across adjacent ages.

We found that the effects of family economic resources on children's outcomes at 14, 24, and 36 months were completely mediated by parenting quality at each age. Income per capita in the family and the other family resources variables had statistically significant indirect effects on child outcomes at every age. Although the effects of family income and other family resources on parenting quality were small, they remained unique predictors even after controlling for prior parenting quality. In this investigation of early development, economic resources influenced children primarily through parents' behaviors, which accords with the finding that socioeconomic status indirectly affects children's performance on language outcomes and academic achievement through its influence on parents' beliefs and behaviors (Davis-Kean, 2005; Raviv, Kessenich, & Morrison, 2004). It is likely that economic resources also influence children's development through other paths, including child-care settings, schools, peers, and neighbors, as children age.

Parenting quality also had unique effects on current child cognitive outcomes even after considering child outcomes at prior ages. Moreover, the effects of current parenting quality on current child cognitive outcomes were significant even after controlling for parenting quality at previous ages. Therefore, we present an extremely conservative approach to modeling economic resources and parenting influences on child development.

We found that the effects of lagged parenting quality on current child outcomes, net of the effects of lagged child outcomes, were significant. Our results also showed that past child cognitive outcomes significantly influenced current parenting quality. Furthermore, a model with reciprocal effects from children to parents and vice versa provided a better fit to the data than a model that did not include those paths. The presence of significant reciprocal effects, even after considering prior parenting experiences, suggests that although parenting is stable over time, parents continue to modify their behaviors in response to children's developmental achievements. Beyond cognitive development, other characteristics in children, such as temperament and even physical attributes, have been shown to affect parenting over time; thus, children are active participants in the construction of their own experiences.

Finally, these findings support a cumulative model of developmental influence. The estimated model resulted in increasing explained variance of the Bayley MDI over successive child ages. By 36 months, 40% of the variance in Bayley MDI was accounted for by model estimates that included prior and current measures of parenting, family resources, and child development. Thus, although specific influences on children resulted in relatively small increments on cognitive achievement at any given age when single variables were considered, these small effects accumulated over time to explain a substantial portion of the variance in cognitive development by the time children reached age 3.

The present study should be interpreted in light of certain limitations. The measures of family resources that were considered do not reflect all dimensions of the stock of human capital in a family nor did analyses include all types of parenting behaviors. For instance, parents' time allocation and their goals for their children's achievement could influence both parenting quality and children's cognitive development. In addition, in constructing our measure of income per capita, we assumed that each member of the family receives an equal share of the family income. A better measure of the resources (including income) that the family allocates to the child would have been the expenditures of the family on child-specific goods (Meyer & Sullivan, 2003). However, this information was not available from our data. Researchers have found positive associations between expenditures on child-specific goods and cognitive development outcomes of infants from low-income and ethnically diverse backgrounds (Lugo-Gil, Yoshikawa, & Tamis-LeMonda, 2007). Moreover, the effect size of family resources on children's cognitive performance reported here is likely a low estimate given our sample characteristics. The current sample was restricted to low-income families, which resulted in a relatively narrow range of family resources compared to the population at large. It is therefore even notable that family resources still related to parenting quality at all ages after controlling for a range of child and parent characteristics. An important direction for future research would be to consider the joint influences of parenting behaviors and parents' expenditures on children's development outcomes in samples that include participants from a greater range of socioeconomic backgrounds.

Another potential limitation is the creation of composite scores of parenting quality at each age. Specifically, the measure of "parenting quality" is derived from free play interactions, the HOME, and a challenge task, and each of these measures in turn comprise multiple indicators. For example, mothers' positive regard, cognitive stimulation, and sensitivity were coded from the videotaped free play assessments and were combined into a composite of "supportiveness" following precedent as well as based on their high collinearity. On the positive side, the combination of parenting measures and tasks into composite scores that are based in multiple measures with strong reliability coefficients (as here) provides a more robust and comprehensive measurement of parenting than a single item. More practically, it would be unwieldy to present separate model solutions for the multiple indicators of parenting obtained in this study. Nonetheless, the decision to use composite measures precludes examination of whether certain aspects of parenting are more predictive of child outcomes than are others. Research on parenting reveals that different forms or aspects of parenting differentially predict children's developmental outcomes, supporting a specificity model of effects (e.g., Baumwell, Tamis-LeMonda, & Bornstein, 1997; Bornstein & Tamis-LeMonda, 2001; Bornstein, Tamis-LeMonda, & Haynes, 1999; Tamis-LeMonda et al., 2001). However, in the present study, models testing the separate scores of parenting did not vary from those yielded in the overall model. Therefore, we felt confident that the use of a parenting quality composite did not conceal differential contributions of individual items or tasks.

Another potential limitations is the use of Likert scales and categorical data for a number of variables included in models, including low birth weight (dichotomous variable) and mothers' own reading practices (Likert scale with responses that were not equidistant). Although these variables predicted parenting and/or child development, they may have underestimated effects given that categorical measures do not reflect the full range of variation obtained with interval data.

Additionally, in testing our conceptual model, we controlled for parents' ethnicity and race in models. Not shown is the finding that when associations were examined separately in White, Latino, and Black families, patterns of covariation among measures did not vary. However, despite general similarities of influence, parenting is clearly shaped by parents' "ethnotheories" or views about children's development and appropriate ways to parent. Parents from different cultural communities vary in their beliefs and practices, and these differences are evident in the daily routines and experiences of young children (Tamis-LeMonda, 2003; Weisner, 2002). Therefore, future research should attend to both similarities and differences in the nature and effects of socialization processes on children's developmental outcomes across different racial and ethnic populations.

Further studies are also needed on the processes that may determine family income, maternal educational achievement, and other family resources. Although focus here was on the ways that family resources might affect parenting and child development, it is also plausible that parenting behaviors and child development influence the amount of available resources in a family. Future research should consider multiple influences among parenting behaviors, children's development outcomes, and family resources within a framework in which the child is viewed as an investment that yields returns across the span of the child's lifetime (Brown & Flinn, 2006).

A final caveat of the current paper is that the statistical approach to modeling reciprocal influences between parenting quality and children's cognitive development did not fully account for simultaneity biases in our estimates. That is, although we analyzed reciprocal effects between parenting quality and child cognitive development outcomes, one could argue that these processes are determined at the same time. Evidence from level and change models of parenting quality and child cognitive development suggests that concerns about selection bias should not be critical. However, to fully address both the selection bias and the issue of simultaneity, we should develop and estimate a model of parents' behavior in which parenting quality and outcomes for children are determined simultaneously (Dahl & Lochner, 2005; NICHD Early Child Care Research Network & Duncan, 2003). This was beyond the scope of this article and it is a matter of consideration for future research.

We believe that our results have important policy implications in light of the finding that family income effects were relatively small and mediated by parenting quality. First, not unlike previous research in this area (Yeung et al., 2002), our results imply that programs aiming solely at supplementing family earnings may not have a strong impact on child cognitive development. In contrast, programs that offer a combination of cash assistance and services designed to improve the quality of parenting may be more effective in promoting cognitive development outcomes among low-income children during the important years that precede entry to school.

Second, our results support services focused on enhancing parents' resources beyond income alone. Strengthening the quality of parenting should also include services aimed at improving family literacy and education, reducing parental stress, and providing high-quality child care. Of all measures in the model, mothers' WJ scores were the strongest predictors of parenting and mothers' education strongly predicted both parenting quality and children's Bayley MDI scores.

In closing, the resources that parents bring to their families satisfy a number of needs in the family, one of them being support for the development of children. Thus, supplementing family income or providing parents with employment and child care services might not be enough to improve low – income children's development outcomes (Morris, Duncan, & Rodrigues, 2007). Children's participation in preschool education intervention programs and efforts to support positive parenting in low-income households are also vitally important to preparing children for later entry to school.

Notes

1. Detailed information on the instruments used in the EHS study, on how the data were collected, and on other studies based on the EHS database can be found in Administration of Children and Families, Office of Planning, Research, and Evaluation (2007) and Mathematica Policy Research Inc. (2007).
2. Among the children included in our sample (N = 2,089), those who had missing data on child outcomes (Bayley, 1993) were more likely to be male, to have mothers with lower levels of education and income at baseline, and to be living in the Midwest and less likely to be living in two-parent households. Mothers who had missing data on parenting quality at any age were more likely to be living alone at baseline, to be Black, to be less educated, to have lower income levels at baseline, and to be living in the Midwest. Participants who had missing data on income per capita were more likely to be mothers of firstborns, to be Black, to be living with their mother and other adults at baseline, to be younger, to have lower levels of education and income at baseline, and to be living in the Northeast and the Midwest.
3. The norming sample of the BSID – III was a national, stratified random sample of 1,700 children aged 1 – 42 months. For the MDI, the average Cronbach's alpha across age groups was .88 and interrater reliability was .96.
4. We also estimated alternative specifications of the model in which the three parenting constructs (supportiveness, quality of the home environment, and sensitivity and regard in challenge tasks) were included separately. We found that each parenting task/assessment contributes to children's cognitive development at 14, 24, and 36 months to a similar magnitude. In addition, these model specifications did not adjust to the data significantly better than the specification that used the parenting quality composite. Therefore, we concluded that the use of a parenting quality composite did not conceal differential contributions of the individual tasks. Results from estimations of alternative specifications of the model are available from the authors upon request.
5. Estimates of the disturbances of parenting quality and child outcomes, of the unanalyzed associations (or observed correlations) between the family resources variables, and of the paths between background factors and parenting are available from the authors upon request.
6. The reported statistical significance of the estimates of indirect effects presented in Table 8 is based on bootstrap standard errors, obtained from the bootstrap option available in MPlus.

References

Ackerman, B., Brown, E., & Izard, C. (2004). The relations between contextual risk, earned income, and the school adjustment of children from economically disadvantaged families. *Developmental Psychology, 40,* 204 – 216.

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.

Administration of Children and Families, Office of Planning, Research, and Evaluation. (2007). *Early Head Start Research and Evaluation Project, 1996-current*. Retrieved May 4, 2007, from http://www.acf.hhs.gov/programs/opre/ehs/ehs_resrch/

Ainsworth, M., Blehar, M. C., Waters, E., & Wall, S. (1978). *Patterns of attachment*. Hillsdale, NJ: Erlbaum.

Aiyagari, S. R., Greenwood, J., & Seshadri, A. (1999). *Efficient investment in children* (Discussion paper 132). Federal Reserve Bank of Minneapolis, MN.

Altonji, J. G., Elder, T. E., & Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of Political Economy, 113,* 151 – 184.

Barnard, K. E. (1976). *NCAST II learners resource manual*. Seattle, WA: NCAST.

Baumrind, D. (1973). The development of instrumental competence through socialization. In A. D. Pick (Ed.), *Minnesota symposium on child psychology* (Vol. 7, pp. 3 – 46). Minneapolis: University of Minnesota Press.

Baumwell, L., Tamis-LeMonda, C. S., & Bornstein, M. H. (1997). Maternal verbal sensitivity and child language comprehension. *Infant Behavior & Development, 20,* 247 – 258.

Bayley, N. (1993). *Manual for Bayley Scales of Infant Development* (2nd ed.). San Antonio, TX: Psychological Corporation.

Becker, G. (1964). *Human capital: A theoretical and empirical analysis with special reference to education*. Cambridge, MA: National Bureau of Economic Research.

Becker, G. (1991). *A treatise on the family: Enlarged edition*. Cambridge, MA: Harvard University Press.

Bell, R. (1968). A reinterpretation of the direction of effects in studies of socialization. *Psychological review, 75*, 81–95.

Berlin, L. J., Brady-Smith, C., & Brooks-Gunn, J. (2002). Links between childbearing age and observed maternal behaviors with 14-month-olds in the Early Head Start Research and Evaluation Project. *Infant Mental Health Journal, Special Issue: Early Head Start, 23*, 104–129.

Blau, D. (1999). The effects of income on child development. *Review of Economics and Statistics, 81*, 261–276.

Bloom, L. (1993). *The transition from infancy to language*. Cambridge, MA: Cambridge University Press.

Bornstein, M. H. (2002). Parenting infants. In M. H. Bornstein (Ed.), *Handbook of parenting* (Vol. 1, 2nd ed., pp. 3–44). Mahwah, NJ: Erlbaum.

Bornstein, M. H., & Tamis-LeMonda, C. S. (2001). Mother-infant interaction. In G. Bremner & A. Fogel (Eds.), *Blackwell handbook of infant development. Handbook of developmental psychology* (pp. 269–295). Malden, MA: Blackwell.

Bornstein, M. H., Tamis-LeMonda, C. S., & Haynes, M. O. (1999). First words in the second year: Continuity, stability, and models of concurrent and predictive correspondence in vocabulary and verbal responsiveness across age and context. *Infant Behavior & Development, 22*, 65–85.

Bradley, R., Caldwell, B. M., Rock, S. L., Ramey, C. T., Barnard, K. E., Carol, G., et al. (1989). Home environment and cognitive development in the first 3 years of life: A collaborative study involving six sites and three ethnic groups in North America. *Developmental Psychology, 25*, 217–235.

Bradley, R., Whiteside, L., Mundfrom, D. J., & Blevins-Knabe, B. (1995). Home environment and adaptive social behavior among premature, low birth weight children: Alternative models of environmental action. *Journal of Pediatric Psychology, 20*, 347–362.

Brooks-Gunn, J., & Chase-Lansdale, P. L. (2002). Adolescent parenthood. In M. H. Bornstein (Ed.), *Handbook of parenting* (Vol. 3, 2nd ed., pp. 173–214). Mahwah, NJ: Erlbaum.

Brooks-Gunn, J., & Duncan, G. J. (1997). The effects of poverty on children. *Future of Children, 7*, 55–71.

Brown, M., & Flinn, C. J. (2006). *Investment in child quality over marital states*. Unpublished manuscript, University of Wisconsin at Madison and New York University.

Bruner, J. (1983). Child's talk: Learning to use language. New York: Norton.

Caldwell, B. M., & Bradley, R. H. (1984). *Home observation for measurement of the environment*. Little Rock: University of Arkansas at Little Rock.

Campbell, F. A., & Ramey, C. T. (1994). Effects of early intervention on intellectual and academic achievement: A follow-up study of children from low-income families. *Child Development, 65*, 684–698.

Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social cognition, joint attention and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development, 63*(Serial No. 255).

Collins, A. W., Maccoby, E. E., Steinberg, L., Hetherington, E. M., & Bornstein, M. H. (2000). Contemporary research on parenting: The case for nature and nurture. *American Psychologist, 55*, 218–232.

Dahl, G. B., & Lochner, L. (2005). *The impact of family income on child achievement* (Working Paper No. 11279). Cambridge, MA: NBER Working Paper Series.

Davies, J. B., & Zhang, J. (1995). Gender bias, investments in children, and bequests. *International Economic Review, 36*, 795–818.

Davis-Kean, P. E. (2005). The influence of parent education and family income on child achievement: The indirect role of parental expectations and the home environment. *Journal of Family Psychology, 19*, 294–304.

Duncan, G. J., Brooks-Gunn, J., & Klebanov, P. K. (1994). Economic deprivation and early childhood development. *Child Development, 65*, 296–318.

Duncan, G. J., & Magnuson, K. A. (2005). Can family socioeconomic resources account for racial and ethnic test score gaps? *Future of Children, 15*(1), 35–54.

Duncan, G. J., Yeung, J. W., Brooks-Gunn, J., & Smith, J. R. (1998). How much does childhood poverty affect the life chances of children? *American Sociological Review, 63*, 406–423.

Ermisch, J., & Francesconi, M. (2000). Educational choice, families, and young people's earnings. *Journal of Human Resources, 35*, 143–176.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., & Pethick, S. J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59*(Serial No. 242).

Foster, E. M. (2002). How economists think about family resources and child development. *Child Development, 73*, 1904–1914.

Garrett, P., Ng'andu, N., & Ferron, J. (1994). Poverty experiences of young children and the quality of their home environments. *Child Development, 65*, 331–345.

Guo, G., & Harris, K. M. (2000). The mechanisms mediating the effects of poverty on children's intellectual development. *Demography, 37*, 431–447.

Hart, B., & Risley, T. (1995). *Meaningful differences in the everyday experiences of young American children*. Baltimore: Paul H. Brookes.

Haveman, R., & Wolfe, B. (1995). The determinants of children's attainments: A review of methods and findings. *Journal of Economic Literature, 33*, 1829–1878.

Hill, M. A., & O'Neill, J. (1994). Family endowments and the achievement of young children with special reference to the underclass. *Journal of Human Resources, Special Issue: The Family and Intergenerational Relations, 29*, 1064–1100.

Hirsh-Pasek, K., & Golinkoff, R. (2002) Language development. In N. Salkind (Ed.), *Child development* (pp. 227–232). New York: Macmillan.

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development, 74,* 1368 – 1378.

Hoff, E., Laursen, B., & Tardif, T. (2002). Socioeconomic status and parenting. In M. H. Bornstein (Ed.), *Handbook of parenting* (Vol. 2, 2nd ed., pp. 231 – 252). Mahwah, NJ: Erlbaum.

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology, 27,* 236 – 248.

Ispa, J. M., Fine, M. A., Halgunseth, L. C., Harper, S., Robinson, J., Boyce, L., et al. (2004). Maternal intrusiveness, maternal warmth, and mother-toddler relationship outcomes: Variations across low-income ethnic and acculturation groups. *Child Development, 75,* 1613 – 1631.

Keown, L. J., Woodward, L. J., & Field, J. (2001). Language development of pre-school children born to teenage mothers. *Infant and Child Development, 10,* 129 – 145.

Landry, S. H., Smith, K. E., & Swank, P. R. (2006). Responsive parenting: Establishing early foundations for social, communication, and independent problem-solving skills. *Developmental Psychology, 42,* 627 – 642.

Landry, S. H., Smith, K. E., Swank, P. R., Assel, M. A., & Vellet, S. (2001). Does early responsive parenting have a special importance for children's development or is consistency across early childhood necessary? *Developmental Psychology, 37,* 387 – 403.

Leseman, P. P. M., & de Jong, P. F. (1998). Home literacy: Opportunity, instruction, cooperation, and social-emotional quality predicting early reading achievement. *Reading Research Quarterly, 33,* 294 – 318.

Levy, D., & Duncan, G. J. (2000). *Using sibling samples to assess the effects of childhood family income on completed schooling* (Working paper no. 168). Evanston, IL: Joint Center for Policy Research.

Love, J. M., Kisker, E. E., Ross, C., Raikes, H., Constantine, J., Boller, K., et al. (2005). The effectiveness of Early Head Start for 3-year-old children and their parents: Lessons for policy and programs. *Developmental Psychology, 41,* 885 – 901.

Lugo-Gil, J., Yoshikawa, H., & Tamis-LeMonda, C. S. (2007, April). *Associations of child-specific expenditures and parental support of learning with the cognitive development of ethnically diverse and immigrant infants.* Paper presented at the Society for Research in Child Development Biennial Meeting, Boston.

Maccoby, E. E., & Martin, J. A. (1983). Socialization in the context of the family: Parent-child interaction. In P. H. Mussen (Series Ed.) & E. M. Hetherington (Vol. Ed.), *Handbook of child psychology: Vol. 4. Socialization, personality, and social development* (4th ed., pp. 1 – 101). New York: Wiley.

Magnuson, K. (2005, April). *Does early childhood behavior affect achievement in middle childhood and early adolescence?* Paper presented at the Society for Research in Child Development Biennial Meeting, Atlanta, GA.

Matas, L., Ahrend, R. A., & Sroufe, L. A. (1978). The continuity of adaptation in the second year: Relation between quality of attachment and later competence. *Child Development, 49,* 547 – 556.

Mathematica Policy Research, Inc. (2007). *Early Head Start research and evaluation.* Retrieved May 4, 2007, from http://www.mathematica-mpr.com/earlycare/ehstoc.asp.

Mayer, S. E. (1997). *What money can't buy: Family income and children's life chances.* Cambridge, MA: Harvard University Press.

McCartney, K., Bub, K. L., & Burchinal, M. (2006). Selection, detection and reflection. *Monographs of the Society for Research in Child Development, 71*(Serial No. 285).

Melzi, G., Paratore, J. R., & Krol-Sinclair, B. (2000). Reading and writing in the daily lives of Latino mothers participating in an intergenerational project. *National Reading Conference, 49,* 178 – 193.

Meyer, B. D., & Sullivan, J. X. (2003). Measuring the well-being of the poor using income and consumption. *Journal of Human Resources, 38*(Suppl. 5), 1180 – 1220.

Moore, K. A., Evans, V. J., Brooks-Gunn, J., & Roth, J. (2001). What are good child outcomes? In A. Thorton (Ed.), *The well-being of children and families, research and data needs* (pp. 59 – 84). Ann Arbor: The University of Michigan Press.

Morris, P., Duncan, G. J., & Rodrigues, C. (2007). *Does money really matter? Estimating impacts of family income on young children's achievement with data from random-assignment experiments.* Manuscript submitted for publication.

Muthén & Muthén. (2004). *Mplus: Statistical analysis with latent variables* (Version 3) [Computer software]. Los Angeles, CA: Author.

National Institute of Child Health and Human Development Early Child Care Research Network. (2001). Child care and family predictors of preschool attachment and stability from infancy. *Developmental Psychology, 37,* 847 – 862.

National Institute of Child Health and Human Development Early Child Care Research Network. (2005). Duration and developmental timing of poverty and children's cognitive and social development from birth through third grade. *Child Development, 76,* 795 – 810.

National Institute of Child Health and Human Development Early Child Care Research Network & Duncan, G. J. (2003). Modeling the impacts of child care quality on children's preschool cognitive development. *Child Development, 74,* 1454 – 1475.

Nelson, K. (1973). Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development, 38*(Serial No. 149).

Nelson, K. (1988). Constraints on word learning? *Cognitive Development, 3,* 221 – 246.

Paratore, J. R., Melzi, G., & Krol-Sinclair, B. (2003). Learning about the literate lives of Latino mothers. In D. M. Barone & L. M Morrow (Eds.), *Literacy and young children: Research-based practices in early literacy* (pp. 101 – 118). New York: Guilford Press.

Pettit, G. S., Bates, J. E., & Dodge, K. A. (1997). Supportive parenting, ecological context, and children's adjustment: A seven-year longitudinal study. *Child Development, 68,* 908–923.

Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1,* 385–401.

Raikes, H., Luze, G., Brooks-Gunn, J., Raikes, H. A., Pan, B. A., Tamis-LeMonda, C. S., et al. (2006). Mother–child book reading in low-income families: Correlates and outcomes during the first three years of life. *Child Development, 77,* 924–953.

Raviv, T., Kessenich, M., & Morrison, F. (2004). A mediational model of the association between socioeconomic status and three-year-old language abilities: The role of parenting factors. *Early Childhood Research Quarterly, 19,* 528–547.

Rodriguez, E. T., Tamis-LeMonda, C. S., & Spellmann, M. (2005, April). *Children's early literacy environment: Promoting language and cognitive development over the first four years of life.* Paper presented at the Society for Research in Child Development Biennial Meeting, Atlanta, GA.

Rohner, R. P. (1986). *The warmth dimension: Foundation of parental acceptance-rejection theory.* Beverly Hills, CA: Sage.

Sameroff, A. J., & MacKenzie, M. J. (2003). Research strategies for capturing transactional models of development: The limits of the possible. *Development and Psychopathology, 15,* 613–640.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7,* 147–177.

Schultz, T. P. (1993). Investments in the schooling and health of women and men: Quantities and returns. *Journal of Human Resources, Special Issue: Symposium on Investments in Women's Human Capital and Development, 28,* 694–734.

Schultz, T. W. (1973). The value of children: An economic perspective. *Journal of Political Economy, Part 2: New Economic Approaches to Fertility, 81,* S2–S13.

Schumacker, R. E., & Lomax, R. G. (1996). *A beginner's guide to structural equation modeling.* Mahwah, NJ: Lawrence Erlbaum.

Shears, J., & Robinson, J. (2005). Fathering attitudes and practices: Influences on children's development. *Child Care in Practice, 11,* 63–79.

Shenkin, S. D., Starr, J. M., & Deary, I. J. (2004). Birth weight and cognitive ability in childhood: A systematic review. *Psychological Bulletin, 130,* 989–1013.

Skinner, E. A., (1986). The origins of young children's perceived control: Mother contingent and sensitive behavior. *International Journal of Behavioral Development, 9,* 359–382.

Snow, C. E. (1986). Conversations with children. In P. Fletcher & M. Garman (Eds.), *Language acquisition* (2nd ed., pp. 69–89). Cambridge, UK: Cambridge University Press.

Storch, S. A., & Whitehurst, G. J. (2001). The role of family and home in literacy development of children from low-income backgrounds. *New Directions for Child and Adolescent Development, 92,* 53–71.

Tamis-LeMonda, C. S. (2003). Cultural perspectives on the 'whats?' and 'whys?' of parenting. *Human Development, 46,* 319–327.

Tamis-LeMonda, C. S., Bornstein, M. H., & Baumwell, L. (2001). Maternal responsiveness and children's achievement of language milestones. *Child Development, 72,* 748–767.

Tamis-LeMonda, C. S., Shannon, J. D., Cabrera, N. J., & Lamb, M. E. (2004). Fathers and mothers at play with their 2- and 3-year-olds: Contributions to language and cognitive development. *Child Development, 75,* 1806–1820.

Votruba-Drzal, E. (2003). Income changes and cognitive stimulation in young children's home learning environments. *Journal of Marriage and Family, 65,* 341–355.

Vygotsky, L. (1978). *Mind in society.* Cambridge, MA: Harvard University Press.

Watson, J. S. (1985). Contingency perception in early social development. In T. Field & N. Fox (Eds.), *Social perception in infants* (pp. 157–176). Norwood, NJ: Ablex.

Weisner, T. (2002). Ecocultural understanding of children's developmental pathways. *Human Development, 45,* 275–281.

Woodcock, R. W., & Johnson, M. B. (1989). *Tests of Achievement, Standard Battery.* Chicago: Riverside.

Woodward, A. L., & Markman, E. M. (1998). Early word learning. In D. Kuhn & R. S. Siegler (Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language* (pp. 371–420). New York: Wiley.

Yeung, J. W., Linver, M. R., & Brooks-Gunn, J. (2002). How money matters for young children's development: Parental investment and family processes. *Child Development, 73,* 1861–1879.

Child Development, July/August 2008, Volume 79, Number 4, Pages 1086–1102

Phonological and Semantic Priming in Children With Reading Disability

Rebecca S. Betjemann and Janice M. Keenan
University of Denver

Lexical priming was assessed in children with reading disability (RD) and in age-matched controls (M = 11.5 years), in visual and auditory lexical decision tasks. In the visual task, children with RD were found to have deficits in semantic (SHIP – BOAT), phonological/graphemic (GOAT – BOAT), and combined (FLOAT – BOAT) priming. The same pattern of semantic priming deficits also occurred in auditory lexical decisions, suggesting that the semantic deficits are not confined to reading. Children with RD also showed less priming than reading-age matched controls, suggesting that their priming deficits are not simply due to lower reading level but are due to the reading disability in particular. These semantic deficits may contribute to both the word reading and the comprehension problems seen in children with RD.

Early models of word recognition assumed that the process of identifying a word is completely independent of the process of retrieving a word's meaning (e.g., Forster, 1976). Much like finding the definition of a word in the dictionary only after one has mapped the spelling onto the dictionary entries, so too in these models, only after the correct lexical entry was identified was the semantic information available.

Current models of word recognition, such as the popular triangle model of Plaut, McClelland, Seidenberg, and Patterson (1996), however, consider semantics to also be a player in the identification of single words, along with orthographic and phonological representations. Even nonconnectionist models, such as more recent formulations of the dual-route model (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001), include a semantic component that contributes to word identification. But despite such depictions, semantics has been, as Keenan and Betjemann (2007) recently termed it, the "neglected side of the triangle" in explanations of single-word reading. Semantics is typically invoked only to explain word identification when a word occurs in context. Also, when a child is found to have problems in word identification out of context, the problems are attributed only to deficits in phonological and orthographic processing, not semantics.

Keenan and Betjemann (2007) tried to counter this neglect by reviewing a number of studies that make a case for the involvement of semantics in single word identification both from a theoretical perspective and from empirical evidence. They contend that, at least in the English language, the case for semantic involvement is compelling. In English, orthographic and phonological pathways alone cannot efficiently master all the words because the mappings between these two domains are often inconsistent (so-called irregular words). In models including semantics, activation spreads from orthography and phonology to semantic representations, and these bidirectional connections with semantics facilitate word recognition and make the process more efficient. Thus, when semantic layers were added to connectionist models that initially only included orthography and phonology (Seidenberg & McClelland, 1989), the models were able to perform better on exception word learning, nonword pronunciation, and lexical decision (Plaut et al., 1996). Furthermore, a number of empirical studies show that although the contribution of semantics may not always be required for normal word identification, it is automatically activated nonetheless (Byrne & Shea, 1979; Forster & Hector, 2002; Meyer & Schvaneveldt, 1971; Seidenberg & Tanenhaus, 1979; Swinney, 1979).

If semantics is a significant contributor to word identification, it raises the question as to whether deficits in semantics may be contributing to word decoding problems in children with reading

---

This research was supported by grant NIH HD027802 to Janice M. Keenan through the Colorado Learning Disabilities Research Center. Rebecca S. Betjemann is now at the Institute for Behavioral Genetics at the University of Colorado and was supported in writing this article by NIMH training grant T32 MH016880-25 to J. Hewitt at the University of Colorado. We are grateful to all the participants and their families; to Dick Olson, John DeFries, and Sally Wadsworth for the word decoding and IQ scores on our participants; to Bruce Pennington and Erik Willcutt for the ADHD scores; to all the testers; and to Laura Roth and Amanda Miller for their comments on the manuscript.

Correspondence concerning this article should be addressed to Rebecca S. Betjemann, Institute for Behavioral Genetics, University of Colorado, UCB 447, Boulder, CO 80309-0447. Electronic mail may be sent to betjeman@colorado.edu.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7904-0018

disability (RD). RD is most typically associated with deficits in phonological and/or orthographic representations (Snowling, 2000; Wagner & Torgesen, 1987), not semantic representations; however, Vellutino Scanlon, and Spearing (1995) suggest that whereas phonological deficits may have the strongest influence on reading problems, deficits in semantics likely contribute to some extent, as well. More typically, semantic representations are thought to be a strength in children with RD because they rely more on semantic context to assist them in word identification. This is shown by studies that contrast word identification performance in context versus out of context. Children with RD show larger context effects than typical readers in terms of both speed and accuracy (Ben-Dror, Pollatsek, & Scarpati, 1991; Bruck, 1990; Nation & Snowling, 1998; Perfetti & Roth, 1981; Simpson & Lorsbach, 1987; Simpson, Lorsbach, & Whitehouse, 1983; Stanovich, 1980).

There are three reasons why this finding should not be interpreted as evidence that semantics is a strength for children with RD. One is that they rely more on context than children with normal reading skills because their decoding is weak and they need to use context to compensate. For example, some studies have used sentence primes, where the sentence context allows readers to predict possible target words that have meanings fitting that context. Poor readers who have difficulty reading the target words out of context may be more likely to strategically use the content of the sentence to guess what the word might be (Stanovich, 1980, 1984; see also Ashby, Rayner, & Clifton, 2005; Frost et al., 2005). Thus, improvement with context says more about these children's problems out of context than about their semantic representations.

The other two reasons why improvement in word identification with context should not be regarded as evidence for better semantic representations are measurement issues. Because children with RD read more slowly, they can show larger context effects than typical readers in terms of the size of the *absolute difference* between in- and out-of-context reading time. However, that difference is often eliminated when one controls for baseline differences in speed of responding (Chapman, Chapman, Curran, & Miller, 1994; Nation & Snowling, 1998; Tunmer & Chapman, 1998). Similarly, poor readers show greater improvement in accuracy of word identification when words occur in context than out of context because they identify fewer words correctly out of context than children with normal reading skill. Thus, whereas good readers are often already near ceiling in the out of context condition, children with RD have more room for improvement.

If poor readers' greater improvement in word identification does not imply that semantics is a strength for them, then it is possible that it could be a deficit. Semantic deficits in children with RD are not unreasonable to hypothesize because it makes sense that the same factors that produced weaker phonological and orthographic representations in these children could very well have also produced weaker semantic representations. Furthermore, because children with RD tend to read less than other children (e.g., Perfetti, Marron, & Foltz, 1996; Stanovich, 1986), they would also have less exposure to semantic information, which could further hinder the development of semantic representations and connections (Vellutino, Scanlon, & Tanzman, 1988; Vellutino et al., 1995).

Support for the possibility of deficits in semantic representations in children with RD comes from the finding that they show poor comprehension not just when they are reading but also when they are listening (e.g., Betjemann, Keenan, & Fazendeiro, 2002; Betjemann, Keenan, & Olson, 2003; Catts, Hogan, & Fey, 2003; Mann, Liberman, & Shankweiler, 1980; Mann, Shankweiler, & Smith, 1984; Spring & French, 1990; Tiu, Thompson, & Lewis, 2003). If the deficits in these children's lexical representations were simply in the mappings of orthography onto phonology, then when such mappings are not involved, as in listening, they should not show a deficit. The fact that they do show deficits in listening comprehension suggests that disabled readers may have broader deficits.

Beyond listening comprehension, there are other studies that suggest impoverished semantic processing using explicit measures. For example, studies have shown that poor readers have deficits on measures of semantic categorization and category recall (e.g., Ben-Dror, Bentin, & Frost, 1995; Bjorklund & Bernholtz, 1986; Lovrich, Cheng, Velting, & Kazmerski, 1997; Vellutino et al., 1995, 1998), measures of receptive and expressive semantic abilities (Purvis & Tannock, 1997), and measures of vocabulary (Assink, Soeteman, & Knuijt, 1999; Vargo, Grosser, & Spafford, 1995; Vellutino et al., 1995). Although these studies suggest that children with RD may have explicit semantic deficits in addition to their phonological deficits, it is unclear if these semantic deficits will be present with measures of implicit semantic processing, such as priming.

The purpose of the research presented here was to explore the deficits in the lexical representations of poor readers, including possible deficits in semantic representations, by examining priming in the lexical decision paradigm. Priming offers the advantage of assessing lexical representations implicitly. For example, in many assessments of phonological representations, such as phonological awareness tasks (e.g., Wagner & Torgesen, 1987), poor performance of children with RD due to deficits in their phonological representations is confounded with their ability to understand and respond to instructions that require explicit manipulation of the representations. In contrast, priming assesses lexical representations uncontaminated by these other factors.

The lexical decision task does not require children to read the words aloud. They simply press one button if the stimulus is a word and another if it is not. Thus, children with RD who might recognize a word but have difficulty pronouncing it are not penalized. In addition, by having children make lexical decisions rather than perform a naming task where they would read the words aloud, we can also examine priming in an auditory lexical decision task. Being able to do a comparison of performance across modalities is very important to determine the extent to which deficits are specific to reading. If any priming difficulties that children experience in the visual domain disappear when they are making the decision based on auditory stimuli, then we can conclude that the deficits are specific to reading. On the other hand, if similar priming deficits are present in both auditory and visual modalities, then it suggests the deficits are in the lexicon and not specific to reading.

Facilitation from priming is evaluated by determining if the time to process a target word is shorter following a related prime word than following an unrelated prime word. If so, the prime is said to have preactivated the target, and thus the target requires less processing to bring it to full activation and identification. Because the amount of facilitation provided by the prime is related to the similarity of prime–target representations, the amount of facilitation can be used to assess the quality of lexical representations and compare them across typical and disabled readers. If a child has deficiencies in their phonological or semantic representations, then we would expect that when the target word is presented, it will not overlap with the prime word as well as it does for children whose representations are not affected, and thus, they will show less facilitation from the prime.

Previous studies of priming in children with RD have often used sentences as primes (Ben-Dror, Pollatsek, & Scarpati, 1991; Bruck, 1990; Nation & Snowling, 1998). As described above, when sentences are used as primes, any differences found between children with RD and skilled readers could simply reflect differences in their use of strategy to predict the target from the prime, rather than differences in the quality of lexical representations. To investigate the quality of lexical representations uncontaminated by possible strategy differences, the current study uses single-word primes.

There have been only a handful of studies that have examined priming in poor readers using single-word primes. One was by Simpson et al. (1983) in which they investigated semantic priming in latencies to read the words aloud rather than to make lexical decisions. They found greater priming in the poor readers than in the good readers. However, it is difficult to interpret these results because they are not adjusted for differences between the groups in baseline response speed, so priming effects are confounded with group differences in speed and accuracy.

A study by Merrill, Sperber, and McCauley (1980) and a study by Assink, Van Bergen, Van Teeseling, and Knuijt (2004) both found no differences in semantic priming between above and below average readers with single-word primes. However, both studies had small sample sizes and thus recommended caution in interpreting their nonsignificant effects.

Booth, Perfetti, and MacWhinney (1999) also used single-word primes but their focus was to investigate phonological and orthographic priming in children with RD. They presented visual primes and targets for very fast, fixed durations for all children, and then the children were asked to write down the target word they saw. Although they found significant deficits in both orthographic and phonological priming in poor readers relative to good readers, there are at least two problems with conclusively interpreting their results as evidence for weaknesses in poor readers' lexical representations. One is that the duration of the prime may not have been sufficient for children with reading problems to identify it or process it fully, as better readers could. When the results showed less priming in poor readers, it could have been because on many trials the exposure was too short for children with RD to process the prime sufficiently. Furthermore, because children had to accurately spell their responses for them to be counted as correct, the results are confounded by spelling deficits that are often seen in children with RD (Snowling, 1998).

The current study attempts to improve on previous research in the following ways. As noted above, we use single-word primes rather than sentences to avoid possible confounds with strategy differences across groups. Second, we avoid the problems of children with RD possibly not identifying the prime by using a procedure called "active priming," which requires that the child make a lexical decision to the prime as well as to the target. Because the prime stays on the screen until the child makes a response, we can be sure that the prime is not presented too fast for the poor readers. Third, we adjust for differences between the poor readers and their controls in baseline speed of responding when we examine the amount of facilitation from related primes. Finally, we offer a broader assessment of lexical representations because we examine priming from multiple types of related primes as we will now explain.

As shown in Table 1, we examine priming from identity primes, phonologically/graphemically related primes, semantic primes that are synonyms, semantic primes that are associates, and what Keenan and colleagues (Keenan, 2000; Keenan, Grubb, & Betjemann, 1999; Keller, 1981) have termed *combined* primes. Combined primes are related to the target both semantically and phonologically/graphemically, such as WINE-DINE. Keenan and colleagues have consistently found that skilled readers' visual lexical decision times show more priming from combined primes than from single-dimension primes (Betjemann & Keenan, 2001; Keenan & Betjemann, 2001, 2002; Keenan et al., 1999; Keller, 1981). This suggests that information from both the semantic and phonological/graphemic similarity is activated and facilitates processing of the target (c.f. Gonnerman, Seidenberg, & Andersen, 2007). Combined primes thus represent a powerful tool for assessing priming in children with RD because if there is a deficit in a single-dimension priming condition such as semantic primes, then we can check the reliability of this finding by determining if it also shows up as a deficit in the combined condition. Thus, the combined prime condition functions as a replication of the single-dimension prime conditions and can provide additional information on potential deficits in children with RD.

The goal of this priming study then is to examine the nature of the lexical representations in children with RD compared to children with normal reading skills. If poor readers have less precise or less efficient semantic, phonological, or orthographic representations than their controls, it should show up as a priming deficit.

Priming deficits can be manifested in two ways. One is within the RD group itself; a priming deficit on a certain dimension would be evident if response time to that prime condition was not significantly different from that of the unrelated condition. Additionally, a priming deficit could be observed in the RD group relative to the control group; such between-group priming deficits would be shown by baseline-adjusted facilitations for a specific type of related prime being significantly less in the RD group compared to that of controls.

It is possible that there can be group differences between children with RD and controls not only on reading ability but also on IQ, vocabulary, and symptoms of attention deficit hyperactivity disorder (ADHD). Thus, we examine group differences on these variables, and then in our priming analyses, we examine whether these variables are related to group differences in priming between children with RD and their controls.

Experiment 1a: Visual Lexical Decision Task

Method

Participants. A sample of 280 twins and their siblings, ages 8–18, who were part of a larger study by the Colorado Learning Disabilities Research Center (CLDRC; DeFries et al. 1997; Olson, 2004) participated in the visual lexical decision task. Participants were identified from school records from 27 different local school districts. Children were invited to participate if they were identified by school records or parent report to have a school history of reading problems or ADHD, and a control group of twin pairs was also invited to participate. The overall sample was 48% male and 52% female. Similar to the local population, 92% of the sample was Caucasian, 4% Hispanic, and other groups made up 1% or less. Fifty-eight percent of the parents had a college education or higher.

From this sample, 78 children with RD were matched with 78 non-RD control children on chronological age (NRD-CA group). The children with RD were defined by their performance on the Timed

Table 1
Examples of Stimuli for Each Prime Condition

Prime condition	Prime – target		Prime – target		Prime – target		Prime – target	
Identity	BOAT	–BOAT	SHIRT	–SHIRT	BAKE	–BAKE	DAY	–DAY
Combined	FLOAT		SKIRT		CAKE		PAY	
Phonological/graphemic	GOAT		FLIRT		RAKE		BAY	
Semantic synonym	SHIP		BLOUSE		ROAST		DAWN	
Semantic associate	SAIL		PANTS		PIE		BIRTH	
Unrelated	YIELD		GLOW		LOSE		MENU	

Table 2

Participant Descriptives for Children With Reading Disability, Chronological-Age Matched Controls (NRD-CA), and Reading-Age Matched Controls (NRD-RA) in Experiments 1a, 1b, and 2

	Experiment 1a: Visual task		Experiment 1b: Visual task with reading-age match			Experiment 2: Auditory task	
	RD ($n = 78$)	NRD-CA ($n = 78$)	RD ($n = 42$)	NRD-RA ($n = 42$)	NRD-CA ($n = 42$)	RD ($n = 78$)	NRD-CA ($n = 78$)
Age	11.5	11.4	13.5	9.6	13.4	11.4	11.5
Decoding composite *z* score	−1.61	0.83	−1.66	0.71	0.78	−1.69	0.80
PIAT Word Recognition raw score			50.6	50.8	66.6		
WISC–R Full Scale IQ	104.7	115.9	103.9	115.5	115.5	103.3	117.1
WISC–R Vocabulary subtest standard score	10.3	12.8	10.0	12.7	12.7	10.0	13.1
ADHD symptoms mean (scale 0–3)	0.90	0.63	0.76	0.70	0.60	0.94	0.71

Note. PIAT = Peabody Individual Achievement Test; WISC–R = Wechsler Intelligence Scale for Children–Revised; ADHD = attention deficit hyperactivity disorder.

Word Recognition Test (Olson, Forsberg, Wise, & Rack, 1994) and the Peabody Individual Achievement Test, Word Recognition subtest (PIAT; Dunn & Markwardt, 1970). Participants were classified as RD if their composite of age-adjusted *z*-scores from the two word decoding tasks fell below −0.67 (lower 25%) and were classified as non-RD controls if their composite z score was above 0 (upper 50%). As shown in Table 2, the groups did not differ on age, $F(1, 154) < 1$, but the difference in their mean decoding z scores was highly significant, $F(1, 154) = 575.83$, $p < .001$. Participants in both groups had Wechsler Intelligence Scale for Children–Revised Full Scale IQ (WISC–R FSIQ; Wechsler, 1974) scores over 80 and had a high level of accuracy in the task (70% or higher correct on both critical trials and all trials). As shown in Table 2, there were group differences in IQ, $F(1, 154) = 45.2$, $p < .001$; oral vocabulary (WISC–R Vocabulary subtest; Wechsler, 1974), $F(1, 154) = 45.2$, $p < .001$); and symptoms of ADHD (parent and teacher report on the Disruptive Behavior Rating Scale; Barkley & Murphy, 1998), $F(1, 154) = 7.54$, $p < .01$. It was these differences that led us to investigate the relationship between priming and these factors, as well.

Materials. The stimuli consisted of 72 critical prime-target word–word pairs, 36 word–nonword pairs, and 36 nonword–nonword pairs, so that there were an equal number of word and pronounceable nonword targets. Each participant saw the same 72 critical target words; what varied across participants was which prime type preceded a particular target.

As shown in the examples in Table 1, there were six prime conditions: unrelated, phonological/graphemic (both phonologically and graphemically related to the target), semantic associate, semantic synonym, combined (both phonologically/graphemically *and* semantically related to the target), and identity primes. Each participant saw 12 primes from each of the six conditions. Assignment of targets to prime condition was counterbalanced across participants by having six different stimuli lists, with each list having a different group of 12 primes from each prime condition paired with the same 72 target words. The power of this design is that all targets are paired with all prime conditions equally across participants, rather than targets always being paired with a specific prime condition. Thus, any differences between individual target words in domains such as neighborhood size, frequency, or regularity should not affect any particular prime condition more than another because the targets are the same across all prime conditions.

All items were generated to meet a specific list of criteria.[1] The type of semantic relation between the combined primes and the targets was controlled so that half of the pairs in the combined condition were semantic synonyms and half were semantic associates. There were also 64 practice pairs, created to be similar to the critical stimuli.

Design. The design was a 2 (reading ability group: RD, NRD-CA) × 6 (prime condition: identity, combined, phonological/graphemic, semantic synonym, semantic associate, unrelated) factorial. Each participant received all six prime conditions, so prime condition was a within-subjects factor, and reading ability was a between-subjects factor. The dependent variables were errors in identifying the targets as words and reaction times (RTs) to targets on correct responses.

Procedure. The task was run using the computer program E-Prime Version 1.1. On each trial, a prime and target were presented. Each participant responded to a total of 208 trials in six blocks. The first two blocks of 24 trials each were practice trials. The last four blocks of 40 trials each contained the critical trials. In each of these blocks, the first 4 trials were practice items, and the next 36 trials were critical pairs or nonword target pairs. Critical and nonword stimuli pairs were presented in random order within those trials.

The stimuli were presented on the computer screen in black 18-point Courier New font, centered on a white screen background. The task requirements were identical for all trials, regardless of relationship between the prime and the target. At the start of each trial, there was a 500-ms fixation cross (+) shown at the center of the screen to alert the participant that the stimulus was coming. After the cross, the prime was presented at the same center location, and the participant was to decide whether it was a word or not by pressing *yes* or *no* keys labeled on the keyboard. As soon as they responded to the prime stimulus, the prime disappeared, and then a mask of Xs equal to the length of the prime word was immediately shown for 500 ms. At the offset of the mask, the target immediately appeared and stayed on until the participant responded as to whether it was a word. After the response to the target, there was then a 1,000-ms delay before the next trial began. After each block of trials, feedback on average response time and percent correct for that block of trials was presented on the screen. The entire task took about 20 min to complete.

Results

Target items were included in the analyses of RT only if the responses to both the prime and the target were correct. Outliers, defined as data points beyond 2 *SD* from each individual's mean response time for correct answers, and responses faster than 250 ms were removed, resulting in the removal of 508 trials (4.5%). Target *errors* were counted only on those trials for which the response to the prime was correct, and the removal of RT outliers from the errors resulted in the removal of 573 trials (5.1%). The analyses were conducted both treating subjects as the random factor and treating items as the random factor. Because the results were the same for each, to simplify the presentation, only the analyses treating subjects as the random factor are presented.

Overall RTs and errors to visual lexical decisions. Figure 1 presents the RT data for the visual lexical decision task. Not surprisingly, the children with RD took longer ($M = 811$ ms) to determine that a target is a word than their chronological-age matched controls ($M = 734$ ms), $F(1, 154) = 9.14, p < .01$. They also made more errors, 8.1% for RD versus 3.5% for controls, $F(1, 154) = 28.87$, $p < .001$. There was no evidence of a speed – accuracy trade-off because error rates were positively correlated with RTs (RD: $r = .27$; NRD-CA: $r = .14$), allowing us to therefore interpret the RT differences presented in Figure 1.

The RT data in Figure 1 showed a significant main effect of prime condition, $F(5, 770) = 21.7$, $p < .001$. Overall, the identity condition is the fastest in both groups, and the unrelated condition is slowest, and the related prime conditions fall between these two extremes. The Prime Condition × Reading Ability Group interaction was not significant, $F(5, 770) = 1.47$, $p > .05$.

The identity condition was included as a check on a possible ceiling on priming effects; specifically, if the combined condition ended up not significantly faster than the single-dimension phonological or semantic prime conditions, it was important to know if this was because the combined condition was at the ceiling of how fast an individual could respond. Because RT was significantly faster in the identity condition than in the combined condition in both groups: RD, $F(1, 77) = 9.0, p < .01$, and NRD-CA, $F(1, 77) = 5.78, p < .05$, we can conclude that the combined condition was not at

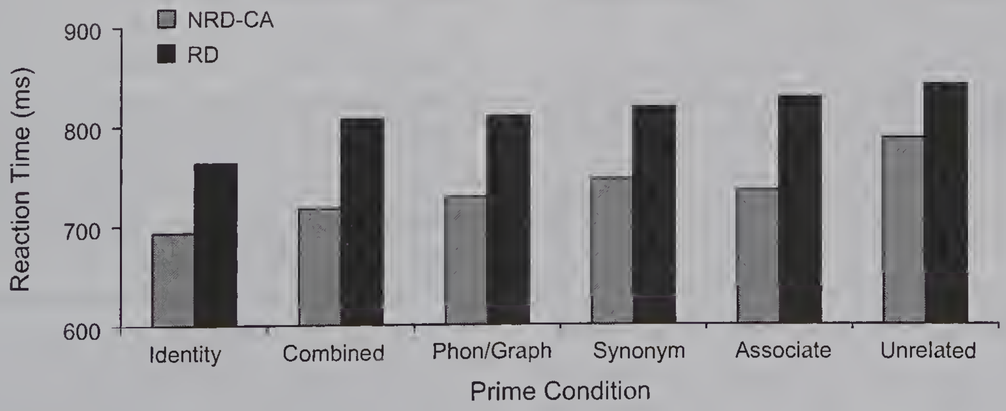


Figure 1. Reaction times in visual task, children with reading disability (RD), and controls (NRD-CA): Experiment 1a.

ceiling. Therefore, the remaining analyses did not include the identity condition.

RTs in the two semantic conditions, the synonym and associate conditions, were not significantly different from one another: RD, $F(1, 77) < 1$, and NRD: $F(1, 11) = 1.73$, $p > .05$. Because combined primes consisted of half semantic synonyms and half semantic associates, we used a mean of these two semantic conditions to compare semantic priming to the other conditions in the analyses of facilitation.

Facilitation from related primes. Facilitation from a related prime occurs if the RT to a target is significantly faster when it is preceded by a related prime than when it is preceded by an unrelated prime. Table 3 presents facilitation values computed as the difference between the mean RT of each prime condition and the mean of the unrelated condition.

For the control group, we found that each of the three related prime types provided significant facilitation in responding to the target. The controls showed significant facilitation in making a lexical decision about a target when it was preceded by a prime that was phonologically/graphemically similar, $F(1, 77) = 28.19$, $p < .001$, or semantically similar, $F(1, 77) = 34.4$, $p < .001$, or similar on both dimensions, combined, $F(1, 77) = 32.06$, $p < .001$. Furthermore, as shown in Table 3, the control group shows a pattern of greater facilitation from a combined prime that activates both semantic and phonological/graphemic dimensions than from either of the single-dimension primes.

The children with RD, like the controls, also showed significant facilitation in the time it takes to classify a target as a word when it was preceded by a word that was phonologically/graphemically related to the target, $F(1, 77) = 5.16$, $p < .05$. Thus, despite their reputed deficits in phonological and orthographic representations, the phonological/graphemic prime activates sufficient lexical information to preactivate related targets and thus producing significant facilitation from the similarity of the prime to the target on these dimensions. The children with RD also show significant facilitation when the prime is from the combined condition, $F(1, 77) = 5.37$, $p < .05$. However, unlike the controls, they do not show significant facilitation when the prime is only related to the target on a semantic dimension, $F(1, 77) = 2.53$, $p = .12$.

In order to directly compare priming between the groups, we need to accommodate the fact that the RD group is overall significantly slower. Chapman et al. (1994) recommend using baseline-adjusted facilitation values. These were computed for each participant in each prime condition by dividing raw facilitation by the participant's mean response time in all word target trials ([unrelated RT – related RT]/mean RT). The RD group showed significantly less facilitation than the NRD-CA control group across each prime condition: combined, $F(1, 154) = 6.43$, $p < .05$; phonological/graphemic, $F(1, 154) = 4.58$, $p < .05$; and semantic, $F(1, 154) = 6.01$, $p < .05$. Also, in contrast to the control group, who showed a tendency to get more facilitation from a combined prime than from a single-dimension prime, the children with RD show no more facilitation from the combined prime than from the phonological/graphemic prime, $F(1, 77) < 1$.

Because there were differences between the RD and the NRD groups in IQ, vocabulary, and symptoms of ADHD, we investigated if these were accounting for the group differences in priming. We computed correlations between baseline-adjusted facilitation in each prime condition with the IQ, vocabulary, and ADHD variables, and found that the correlations were all very low, ranging from −.004 to .16. None of the correlations were significant at the $p < .05$ level, and therefore, when we included these variables as covariates in the analysis, we found that none of the three covariates were significant, as would be expected: FSIQ, $F(1, 151) < 1$; vocabulary, $F(1, 151) < 1$; and ADHD, $F(1, 151) < 1$. Thus, despite overall group differences in these three variables, they are not significantly related to priming, and we therefore present the results without the covariates from here forward.

Table 3
Reaction time Facilitation (ms) in Visual Lexical Decision in Children With Reading Disability (RD) and Controls (NRD-CA): Experiment 1a

	Facilitation		Baseline-adjusted facilitation (facilitation/mean RT)	
	RD	NRD-CA	RD	NRD-CA
Combined (FLOAT – BOAT)	33.8	69.6	0.041	0.097
Phonological/ graphemic (GOAT – BOAT)	31.4	59.2	0.036	0.081
Semantic (SHIP – BOAT)	17.8	46.9	0.022	0.063

Although implicit tasks are believed to be minimally sensitive to strategic responding, over the course of the experiment, participants could potentially adopt strategies that might change the way they respond. These strategy differences could affect one type of prime relationship more than another, or one group of participants more than another; therefore, it is important to rule out the possibility that this is driving our results. We assessed strategy changes over time by examining the patterns of priming in the first and second halves of the experiment. Although the overall response times do get faster in the second half of critical trials compared to the first half, $F(1, 149) = 16.85$, $p < .001$, there is not a significant interaction between experiment half and prime condition, $F(2, 300) < 1$, nor is there a three-way interaction among reading group, half, and prime condition, $F(2, 300) < 1$.

Discussion

Our results show that children with RD get less facilitation from related primes, regardless of the type of relatedness, than their age-matched controls. Because children with RD are known to have problems with phonological processing, less priming than controls in the phonological/graphemic condition is not surprising. What is noteworthy, however, is that they are also showing deficits in semantic priming. Not only are they showing deficits in priming relative to the controls, but they also do not even show significant facilitation from semantic primes compared to the unrelated condition. This semantic priming deficit is also reflected in the combined priming condition; for children with RD, the combined priming condition offers no more facilitation than the phonological/graphemic prime by itself, suggesting that for them, the semantic dimension is not significantly activated in the combined primes.

These priming deficits raise the question of what underlies them. The priming differences are not simply due to IQ, oral vocabulary, or ADHD because these measures are not significantly correlated with priming. So, what is it about the lexical representations and connections in the children with RD that is causing these differences? It is important to remember that the RT differences we observed are all on items that the children got correct. In other words, they are occurring on words that the children know, so the deficits in priming cannot simply be attributed to lack of lexical representations. However, it could be that even though they know the words, the children with RD might not know them as well as the controls because they do after all have a lower reading level. Thus, we turn next to consider whether the differences we observed might be due to their reading level.

If the priming deficits that we observed are due to a developmental delay in the RD group, that is, due to their lower reading level, then if we match them to younger children who have the same reading level but who do not have a reading disability, the differences in performance should go away. On the other hand, if the priming deficits that we observed are due to qualitative differences in the children's representations as a function of their having a reading disability, and not just due to lower reading level, then we might expect that our children with RD would continue to show priming deficits even relative to younger children who have the same reading age. Experiment 1b examines this question by comparing a group of children with RD to both a group of reading-age matched children (NRD-RA) and a chronological-age matched NRD-CA group.

Experiment 1b: Visual Lexical Decision Task With Reading-Age Matched Controls

Method

The requirement of having to derive two different control groups from our pool of participants, one matched on age and one matched on reading level, meant that some of the younger RD children from Experiment 1a could not be included in these analyses because their reading ages were so low that they would have required reading age controls much younger than any of the children in our sample. Thus, three new groups were chosen from the pool of 280 children who had participated in the visual lexical decision task described above. The groups for Experiment 1b were smaller than those for Experiment 1a, with only 42 participants per group, and the RD and NRD-CA participants were over 2 years older than in Experiment 1a. The participants with RD were matched to NRD participants on reading age (NRD-RA) as measured by their raw score on the PIAT Word Recognition subtest (Dunn & Markwardt, 1970), and matched to NRD-CA participants on chronological age. Descriptive statistics on the three groups are presented in the middle column of Table 2. The RD and NRD-RA groups were not different on reading age, $F(1, 82) < 1$, and the RD and NRD-CA groups did not differ in chronological age, $F(1, 82) < 1$. The age-adjusted decoding composite scores of the two control groups, NRD-CA and NRD-RA, did not differ, $F(1, 82) < 1$.

Results and Discussion

Overall error rates did not differ across the three groups, $F(2, 123) = 2.17$, $p > .10$. Errors were positively correlated with RTs ($r = .15$, $N = 756$, $p < .001$), indicating no speed – accuracy trade-off.

The RTs to make correct visual lexical decisions on targets are presented in Figure 2. This figure shows that for every prime condition, the NRD-CA group was the fastest, the children with RD were the next fastest, and the younger children with the same reading age as the children with RD were the slowest to make their lexical decisions. A 3 (reading ability group) × 6 (prime condition) repeated measures analysis of variance showed a significant main effect of reading ability group, $F(2, 123) = 19.66$, $p < .001$; planned comparisons showed that the NRD-CA group ($M = 643$) was significantly faster than the RD group ($M = 703.2$), difference $F(1, 82) = 4.69$, $p < .05$, and the RD group was significantly faster than the younger NRD-RA group ($M = 815.4$), difference $F(1, 82) = 15.2$, $p < .001$.

These differences in overall speed of response necessitate that we examine any group differences there might be in facilitation from related primes by examining baseline-adjusted facilitation, as we did in Experiment 1a. These values are shown in Table 4. Because of the smaller sample size, group differences in facilitation were evaluated collapsed across prime types. Figure 3 shows the baseline-adjusted facilitations for each of the three groups. Even with the smaller sample sizes of this experiment, an interesting pattern of group differences is evident. Specifically, the NRD-CA and NRD-RA groups show very similar amounts of facilitation (NRD-CA: $M = .071$; NRD-RA: $M = .070$, difference $F(1, 82) < 1$, and much greater facilitation than that shown by the RD group ($M = .029$). Therefore, even though the NRD-RA group is much younger, and they responded more slowly to the targets, the proportional amount of facilitation they show from having the target preceded by a related prime is very similar to that of the older controls. However, this is not true of the children with RD. They showed less facilitation than either of the control groups: RD versus NRD-CA, $F(1, 82) = 5.14$, $p < .05$, and RD versus NRD-RA, $F(1, 82) = 3.09$, $p = .08$. The children with RD have the same reading level as the NRD-RA children, and they are almost 4 years older, and yet they are showing reduced priming. This shows that the priming deficits in RD are due to the disability per se, not their reading level. It suggests that there are deficiencies in their lexical representations or networks that limit the facilitation from the prime that even younger children's lexical representations provide.

Finding that priming deficits in RD are due to the disability per se, and not reading level, raises the question of whether priming deficits would also be observed outside the domain of reading. It is reasonable to think that the deficits might be confined to reading. Children with RD have weak orthographic – phonological mappings (Snowling, 2000), and because reading invokes these weak mappings, it could be that the weak orthographic – phonological mappings slow down the whole network and result in reduced priming across the board when reading. If that is the case, then if we have children with RD simply listen to words and make auditory lexical decisions, they should no longer show priming deficits. On the other hand, if the deficits in lexical representations are broader than weak orthographic – phonological mappings, then we might expect to see priming deficits even when reading is not required; that is, the children with RD should show priming deficits in auditory lexical decisions just like they did in the visual lexical decisions. The next experiment uses auditory lexical decision to answer this question.

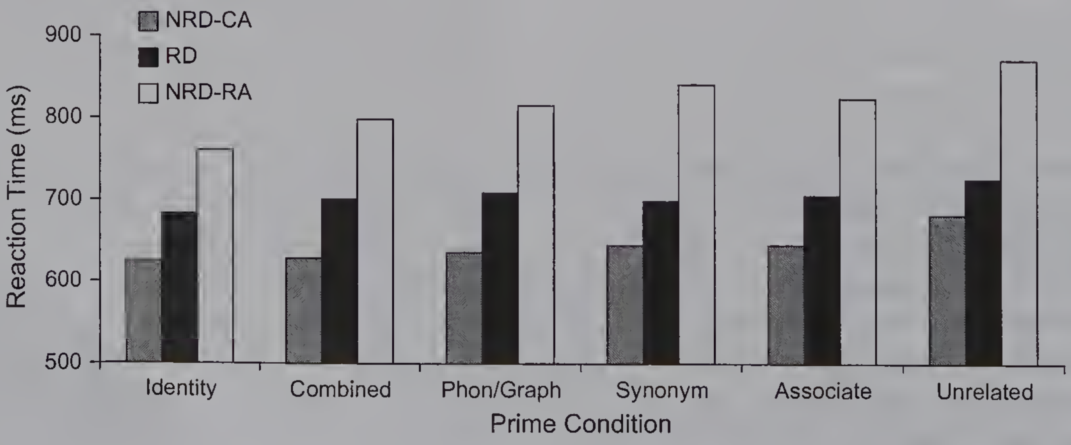


Figure 2. Reaction times in visual task, children with reading disability (RD), reading-age matched controls (NRD-RA), and chronological-age matched controls (NRD-CA): Experiment 1b.

Table 4

Reaction time (RT) Facilitation (ms) in Visual Lexical Decision in Children With Reading Disability (RD), Reading-Age Matched Controls (NRD-RA), and Chronological-Age Matched Controls (NRD-CA): Experiment 1b

	Facilitation			Baseline-adjusted facilitation (facilitation/mean RT)		
	RD	NRD-RA	NRD-CA	RD	NRD-RA	NRD-CA
Combined (FLOAT – BOAT)	24.1	71.9	23.1	0.034	0.090	0.084
Phonological/graphemic (GOAT – BOAT)	16.5	55.4	43.8	0.022	0.069	0.071
Semantic (SHIP – BOAT)	23.1	45.5	36.8	0.032	0.052	0.058

Experiment 2: Auditory Lexical Decision Task

Method

Participants. A new group of 276 participants, who were also participating in the larger CLDRC study, participated in the auditory lexical decision task. From these, a group of 78 children with RD and 78 without RD (NRD-CA group) were selected using the same criteria described in Experiment 1a. As shown in Table 2, the mean ages of the two groups were not significantly different, $F(1, 154) < 1$, but the RD group was significantly impaired compared to the NRD-CA group on word decoding, $F(1, 94) = 656.91$, $p < .001$, as well as on IQ, $F(1, 154) = 67.2$, $p < .001$; WISC vocabulary, $F(1, 154) = 66.8$, $p < .001$; and ADHD symptoms, $F(1, 154) = 4.94$, $p < .05$. We matched the RD and NRD-CA participants in the auditory lexical decision to the participants from these groups in the visual lexical decision task of Experiment 1a; they were matched on both decoding ability and chronological age so that we could do comparisons across modalities.

Materials. Digitally recorded critical word stimuli from the visual task were used in this auditory task.[2]

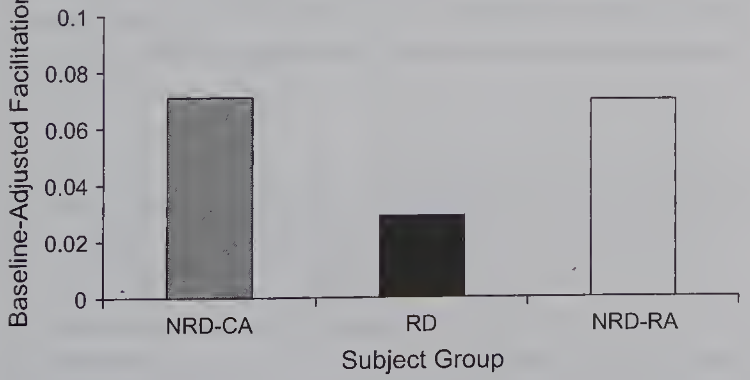


Figure 3. Mean baseline-adjusted facilitation in visual task, children with reading disability (RD), reading-age matched controls (NRD-RA), and chronological-age matched controls (NRD-CA): Experiment 1b.

Design. The design of this experiment is the same as was described in Experiment 1.

Procedure. The procedure in this task was the same as in the visual task in Experiment 1, except for a few differences due to the different modality of presentation. In this task, the stimuli were recorded into digital audio sound files and were played for the participants through headphones. Piloting showed that the children were most attentive if at the start of each trial, an alerting tone was played while simultaneously showing a fixation cross (+) at the center of the screen, that is, giving the child something to look at signaling the imminent occurrence of the prime and target over the headphones. The procedure on each trial thus involved playing an alerting tone at the beginning with the fixation cross on the screen, then after a 500-ms delay, the prime was played; once the child pressed a button signaling whether the sound corresponding to the prime was either a real word or a nonword, then there was a period of 500 ms of silence with a blank screen, and then the target was played and the child pressed a button for the lexical decision on the target. Feedback was given at the same intervals, following the same procedure, as in the visual lexical decision, and was presented visually on the screen.

Results

Overall RTs and errors to auditory lexical decisions. Data were trimmed as described in Experiment 1a, resulting in the removal of outliers in 3.8% of the trials for the RT analyses and 2.9% of the trials in the errors analyses. Figure 4 shows the auditory lexical decision times for the six prime conditions in the RD and NRD-CA groups. Although this was a listening task, not a reading task, the children with RD were nonetheless significantly slower (M =1,190 ms) than the control group (M = 1,096 ms) to determine that the targets that they heard were

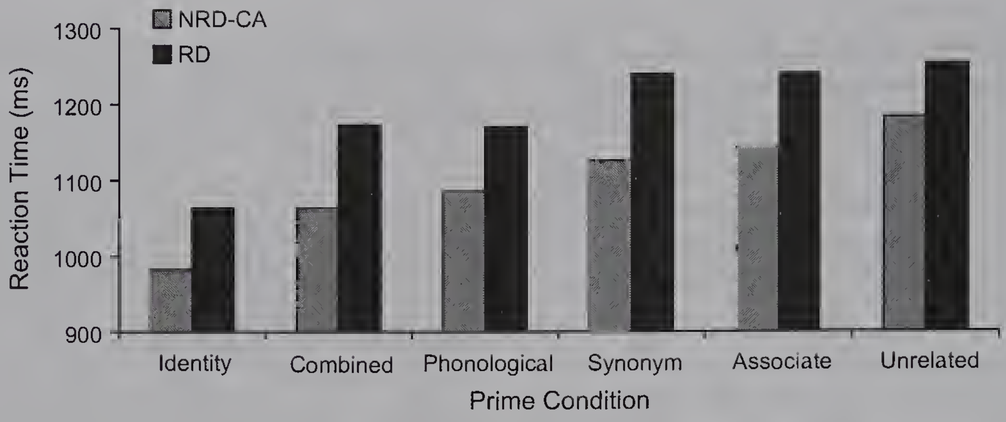


Figure 4. Reaction times in auditory task, children with reading disability (RD), and controls (NRD-CA): Experiment 2.

words, $F(1, 154) = 8.25$, $p < .01$; there was no significant interaction with prime condition $F(5, 770) = 1.11$. The RD group also made significantly more errors on the word targets than the NRD-CA group, 8.4% versus 4.8%; $F(1, 154) = 9.70$, $p < .01$, and again no significant interaction with prime condition, $F(5, 770) < 1$. Error rates and RTs for each prime condition were positively correlated (RD: $r = .25$; NRD-CA: $r = .26$), suggesting no evidence of a speed-accuracy trade-off.

The auditory lexical decision times in Figure 4 also show a significant main effect of prime condition, $F(5, 770) = 77.79$, $p < .001$. Furthermore, it looks very similar to that of the visual lexical decision RTs in Figure 1, in that for both, the identity condition was the fastest and the unrelated condition was the slowest. As in the visual task, we again found that RT in the identity condition was significantly faster than that in the combined condition in both groups: RD, $F(1, 77) = 27.52$, $p < .001$, and NRD-CA, $F(1, 77) = 35.79$, $p < .001$, showing that the combined condition was not at ceiling. RTs in the synonym and associate conditions were not significantly different from one another: RD, $F(1, 77) < 1$ and NRD, $F(1, 77) = 1.91$, $p > .05$; therefore, the semantic mean was used in the analyses of facilitation from related primes.

Facilitation from related primes. Table 5 presents facilitation values computed as the difference in the mean RT of each related prime condition compared to the mean of the unrelated condition. Just as we found in the visual lexical decision task, for the control group, each of the three related prime types provided significant facilitation in responding to the target. The controls showed significantly facilitated RTs to make a lexical decision about an auditory target when it was preceded by a prime that was phonologically similar, $F(1, 77) = 49.07$, $p < .001$, or semantically similar, $F(1, 77) = 16.16$, $p < .001$, or similar on both dimensions (combined), $F(1, 77) = 68.76$, $p < .001$.

The children with RD also showed significant phonological priming, $F(1, 77) = 24.57$, $p < .001$. However, in contrast to the control group, they did not show significant semantic facilitation, $F(1, 77) < 1$. Finding a semantic deficit in the RD group in making auditory lexical decision just like we found in the visual lexical decision suggests that the semantic deficit is not something that only happens when they are reading and invoking weak orthographic–phonological mappings. Instead, it appears rather robust and indicative of qualitative differences in their lexical representations.

Because the RD group had longer RTs than the NRD-CA group, we again computed baseline-adjusted facilitation values for the three prime conditions, as described in Experiment 1a. As in the visual modality, IQ, vocabulary, and ADHD symptoms were not correlated with facilitation (*r*s from −.12 to .07) and

Table 5
Reaction time (RT) Facilitation (ms) in Auditory Lexical Decision in Children With Reading Disability (RD) and Controls (NRD-CA): Experiment 2

	Facilitation		Baseline-adjusted facilitation (facilitation/mean RT)	
	RD	NRD-CA	RD	NRD-CA
Combined (FLOAT–BOAT)	78.1	117.9	0.078	0.109
Phonological (GOAT–BOAT)	81.0	95.4	0.075	0.088
Semantic (SHIP–BOAT)	11.7	46.7	0.019	0.046

were not significant as covariates in the analysis: FSIQ, $F(1, 151) < 1$; vocabulary, $F(1, 151) < 1$; and ADHD, $F(1, 151) < 1$. Although the main effect of subject group did not reach significance, $F(1, 154) = 2.19$, $p = .14$; nonetheless, for the semantic condition, the difference between the RD and NRD-CA groups was marginally significant, $F(1, 154) = 2.85$, $p = .09$.

Similar to what we found with the visual task, in this auditory task, the adjusted facilitation in the combined condition for the NRD-CA group is greater than that in the single-dimension phonological condition, $F(1, 154) = 3.22$, $p = .07$, suggesting that control children are getting some benefit from both the semantic and phonological dimensions in the combined primes. In the RD group, however, the combined facilitation is almost identical to the phonological facilitation, $F(1, 154) < 1$, suggesting that the children with RD are getting facilitation only from the phonological relatedness of the combined primes and not any additional benefit from the semantic relatedness.

Discussion

The goal of this experiment was to determine if the priming deficits that we observed in children with RD on the visual lexical decision task would still occur when the children did not have to read the stimuli but instead just listened to them. The view that the deficit in lexical representations in children with RD is only evident when they read predicts that when they are not reading and are just listening to make their judgments, they should do as well as controls. However, this did not happen. Lexical decisions were less accurate and took longer in the children with RD, even though the task did not involve reading.

Even on the words that the children with RD knew and made correct decisions on, they again showed no semantic priming, in contrast to the controls. In addition, the amount of priming they showed from a prime that combined both phonological and semantic priming was no greater for these children than from a prime that was only phonological, providing further corroboration that they were getting no benefit from the semantic relatedness of those primes. Thus, we can conclude that the lack of semantic priming that we observed in children with RD when they were doing the visual lexical decision task of Experiment 1 was not just a function of the task requiring reading. It suggests that there may be differences within the semantic representations themselves that lead to semantic deficits even in a non-reading task.

We did not see a significant deficit in the RD group in phonological facilitation in this auditory task, as we had seen in the visual task. The difference in phonological priming deficits between the two modalities could be due to the specificity of the deficits in children with RD. The difference between the auditory and the visual lexical decision tasks in the phonological priming condition is that in the visual task the child must first map the graphemes onto the phonemes in order to benefit from the phonological similarity between the prime and the target. Our finding that children with RD show phonological priming deficits in the visual task and not in the auditory task would suggest that these deficits on the visual task may reflect problems in the connections between phonology and orthography rather than within the phonological representations themselves.

Significant phonological priming in children with RD might seem surprising because they are known to have problems with explicit phonological processing (Snowling, 2000; Wagner & Torgesen, 1987), and there is evidence of implicit phonological processing deficits as well (Boada & Pennington, 2006). It should be noted that in our stimuli, the phonological primes rhyme with the targets; that is, the phonological similarity of the prime and target is in the rime of the word, and the onsets are different. This similarity in rhyme is very apparent in the auditory modality, and the patterns of phonological priming that we observed suggest that it was very apparent to children with RD as well as to controls. Our suggestion that rhyme-based priming could be contributing to our failure to find phonological deficits in the auditory task is consistent with that of other studies that show that, by fourth grade, children with RD are often doing almost as well as controls on rhyming tasks (Bowey, Cain, & Ryan, 1992; Bradley & Bryant, 1978). This also fits with Ziegler and Goswami's (2005) notion that during development, the grain size of the phonological representations shifts from larger to smaller: from units the size of words, to onset rimes, to individual phonemes. Although children with RD are slower at developing the smaller grain sizes, their phonological representations at the onset-rime grain size appear to be adequate enough so as to not show a difference in phonological priming compared to the control group with auditory presentation.

Modality Comparisons

We purposely selected the children for the auditory lexical decision task so that they were matched with the RD and NRD-CA participants in the visual task on both chronological age and reading ability, in order that we could do an analysis with modality as a between-subjects factor to assess the differences between the modalities. As shown in Table 2, there were no significant differences between the four groups of subjects in age, $F(3, 308) < 1$. The differences in decoding composite scores between modalities, within reading groups, were also not significant: RD, $F(1, 154) < 1$, NRD-CA and $F(1, 154) < 1$.

Overall RT was slower in the auditory task than in the visual task, $F(1, 310) = 309.56$, $p < .001$. This difference was expected because RTs are measured from the onset of the target. In the auditory task that includes the time it takes to present the target, which in contrast to the visual presentation, is not simultaneous but rather extended over time. This difference in baseline RTs across modalities required that we compare priming across modalities by using baseline-adjusted facilitation scores. It should be noted that by adjusting for baseline differences in modality, the main effect of modality is eliminated, $F(1, 301) = 1.17$.

The view that the deficit in RD only manifests in reading would predict significant interactions between subject group and modality. However, as we have already seen from the results of Experiments 1a and 2, the patterns of priming across the modalities were quite similar. This was confirmed by finding in this analysis that there were no significant interactions between modality and subject group, either overall, $F(1, 308) < 1$, or with prime condition, Prime Condition × Subject Group × Modality interaction, $F(2, 616) < 1$. The RD group showed less overall facilitation than the controls, $F(1, 308) = 8.92$, $p < .01$, and less facilitation in each prime condition: combined, $F(1, 308) = 8.12$, $p < .01$; phonological/graphemic, $F(1, 308) = 4.35$, $p < .05$; and semantic, $F(1, 308) = 8.59$, $p < .01$, but this did not differ across modality, further supporting the argument that priming deficits in children with RD are not confined to reading.

The only significant interaction with modality was that between prime condition and modality, $F(2, 616) = 5.74$, $p < .01$); it reflects a trend for the phonological and for the combined facilitation to be higher in the auditory task than in the visual task: combined, $F(1, 308) = 2.69$, $p = .10$, and phonological/graphemic, $F(1, 308) = 2.92$, $p < .10$. As noted earlier, this is likely due to the phonological rhyming relationship that is more apparent in the auditory task. Note, however, that there is no modality difference in semantic facilitation, $F(1, 308) < 1$.

General Discussion

The goal of this study was to assess the quality of the lexical representations in children with RD by determining whether they show deficits in priming in lexical decision tasks. We looked not only at phonological/graphemic representations, which are usually the focus of research on children with RD, but also at semantics, a somewhat neglected area in reading disability research (Keenan & Betjemann, 2007). We also assessed priming using combined primes that were related to the targets on both phonological and semantic dimensions, as a replication of the single-dimension prime conditions.

We found an overall priming deficit in the children with RD, and a consistent pattern of semantic priming deficits, in particular. The children with RD did not show significant facilitation compared to the unrelated condition with semantic primes, and they showed deficits in baseline-adjusted semantic facilitation compared to control children. These patterns were apparent in the combined prime condition, as well.

The semantic priming deficit occurred not only in the visual priming task but also in the auditory modality where no reading was required. Finding this deficit in both modalities suggests that the impairments in children with RD extend beyond those directly resulting from the mapping of orthography to phonology or beyond simply a lack of resources for processing when there is a strain on this system by having to decode the words (Perfetti, 1985). Instead, the findings in the auditory task suggest that there is a broader deficit. This deficit could be due to either slowness in activation of the semantic representations (e.g., Shanahan et al., 2006) or deficits in the organization within those representations. Although we cannot distinguish these causes in the current study, either leads to less efficiency overall in the semantic system.

We investigated whether the priming deficits shown by children with RD were just due to their lower reading level by comparing their performance to that of a group of children who were matched to them on reading age. We found that even though the reading-age matched control group was on average 4 years younger than the children with RD and slower to make their lexical decisions, they showed greater priming than the children with RD and comparable priming to the chronological-age matched controls. Hence, the priming deficits observed in the children with RD appear to be due to qualitative differences in activation of their representations stemming from their reading disability in particular and not just their lower reading level.

Priming deficits could be due to general slowing related to IQ, impaired executive functions related to ADHD, or having weaker vocabularies. Because vocabulary could be thought of as an explicit measure of semantics (e.g., Vellutino et al., 1995), one might expect that this measure in particular could be responsible for group differences in semantic priming. Like other researchers, we did find significant group differences in these variables (e.g., Pennington, 2006; Stanovich, 2000; Vellutino et al., 1995; Willcutt, Pennington, Olson, Chhabildas, & Hulslander, 2005). However, despite these group differences, we did not find that vocabulary was significantly correlated with baseline-adjusted facilitation in any prime condition, nor were IQ or ADHD symptoms significantly correlated with priming, after adjusting for baseline. This suggests that the priming task was picking up on some unique aspects of lexical and semantic processing that are impaired in the children with reading problems, beyond those that are measured explicitly by vocabulary and IQ. What might those deficits be? Priming deficits likely reflect impaired efficiency of the semantic network, including slower activation of semantics due to the strength of weights of the connections from phonology and orthography or slower settling of semantics because of less efficient coding of semantic representations. Vocabulary and most IQ subtests are not speeded; thus, even though they measure semantic knowledge, they do not account for deficits that are more specific to *efficiency* of semantic processing.

The semantic deficits we found in this study can help explain why many children with RD have comprehension problems not only when they are reading but also when they are listening (Betjemann et al., 2003; Catts et al., 2003; Mann et al., 1980, 1984; Spring & French, 1990; Tiu et al., 2003). Good comprehension requires that the meanings of words presented in a discourse are quickly activated, and they in turn quickly activate related concepts from long-term memory, so that all this semantic information can be easily integrated together into a coherent understanding (Kintsch, 1988, 1998). The semantic deficits observed in this study would slow down that activation and integration, and require more effortful processing for comprehension. Nation and Snowling (1999) found that semantic priming deficits are related to comprehension problems in nondisabled readers; our results suggest that this would apply in children with RD, as well.

In sum, the priming deficits observed in this study show that the problems in children with RD are not restricted to weaknesses in phonological representations and deficits in orthographic-phonological mappings (Snowling, 2000) but extend also into semantics. Our finding that these children show deficits not only when reading words but also when hearing them also supports the notion that their deficit is broader than is typically characterized. The breadth of the deficit can explain why children with reading problems show deficits in listening comprehension as well as in reading comprehension. We hope that these insights into the lexical representations of children with RD can help guide future research on early prevention and remediation, so that we can know best how to keep them from falling further behind.

Notes

1. Items met the following guidelines (from Keller, 1981). (a) Words were restricted to "relatively common" English words, although no minimum frequency of usage was specified. (b) Words with prominent or preferred single pronunciations were required so that phonemic similarity could be specified. (c) Although at least one word of nine letters in length was allowed, shorter words were preferred. (d) Proper names, contractions, and abbreviations were excluded. (e) Rhyming words (phonological/graphemic and combined) were required to share common orthography, so that they would be both phonemically and graphemically related. Hence, pairs such as WEIGHT–LATE and RARE–PAIR were not used. (f) Word pairings could not result from simple prefix substitution (e.g., INCREASE–DECREASE). (g) The separation of compound words into component words for prime–target pairs was avoided where possible (e.g., BACKTRACK into BACK–TRACK). (h) Unrelated primes were chosen randomly from other prime type stimuli lists as much as possible with the provision that no apparent relationship existed within a pair.
2. Two critical prime words were changed for the auditory task because of low frequency. One target word and its six prime words were also changed for this task. Most of the nonword stimuli were the same; the nonwords were originally developed for use with visual presentation, but then in piloting, some were harder to distinguish from real words in auditory presentation; therefore, a total of 38 of 108 critical-block nonword stimuli were modified for this auditory version.

References

Ashby, J., Rayner, K., & Clifton, C., Jr. (2005). Eye movements of highly skilled and average readers: Differential effects of frequency and predictability. *Quarterly Journal of Experimental Psychology, 58A*, 1065–1086.

Assink, E. M. H., Soeteman, W. P., & Knuijt, P. P. N. A. (1999). Word decoding and picture naming in children with a reading disability. *Genetic, Social, & General Psychology Monographs, 125*, 251 – 268.

Assink, E. M. H., Van Bergen, F., Van Teeseling, H., & Knuijt, P. P. N. A. (2004). Semantic priming effects in normal versus poor readers. *Journal of Genetic Psychology, 165*, 67 – 79.

Barkley, R. A., & Murphy, K. (1998). *Attention-deficit hyperactivity disorder: A clinical workbook* (2nd ed.). New York: Guilford Press.

Ben-Dror, I., Bentin, S., & Frost, R. (1995). Semantic, phonologic, and morphologic skills in reading disabled and normal children: Evidence from perception and production of spoken Hebrew. *Reading Research Quarterly, 30*, 876 – 893.

Ben-Dror, I., Pollatsek, A., & Scarpati, S. (1991). Word identification in isolation and in context by college dyslexic students. *Brain & Language, 40*, 471 – 490.

Betjemann, R. S., & Keenan, J. M. (2001, May). *The effects of very fast prime duration on combinatorial priming*. Poster session presented at the meeting of the Midwestern Psychological Association, Chicago.

Betjemann, R. S., Keenan, J. M., & Fazendeiro, T. A. (2002, May). *Reading and listening comprehension in normal readers and children with reading disability.* Talk presented at the annual meeting of the Midwestern Psychological Association, Chicago.

Betjemann, R. S., Keenan, J. M., & Olson, R. K. (2003, June). *Listening comprehension in children with reading disability.* Poster presented at the meeting of the Society for the Scientific Study of Reading, Boulder, CO.

Bjorklund, D. F., & Bernholtz, J. E. (1986). The role of knowledge base in the memory performance of good and poor readers. *Journal of Experimental Child Psychology, 41*, 367 – 393.

Boada, R., & Pennington, B. F. (2006). Deficient implicit phonological representations in children with dyslexia. *Journal of Experimental Child Psychology, 95*, 153 – 193.

Booth, J. R., Perfetti, C. A., & MacWhinney, B. (1999). Quick, automatic, and general activation of orthographic and phonological representations in young readers. *Developmental Psychology, 35*, 3 – 19.

Bowey, J. A., Cain, M. T., & Ryan, S. M. (1992). A reading-level design study of phonological skills underlying fourth-grade children's word reading difficulties. *Child Development, 63*, 999 – 1011.

Bradley, L., & Bryant, P. E. (1978). Difficulties in auditory organisation as a possible cause of reading backwardness. *Nature, 27*, 746 – 747.

Bruck, M. (1990). Word-recognition skills of adults with childhood diagnoses of dyslexia. *Developmental Psychology, 26*, 439 – 454.

Byrne, B., & Shea, P. (1979). Semantic and phonetic memory codes in beginning readers. *Memory & Cognition, 7*, 333 – 338.

Catts, H. W., Hogan, T. P., & Fey, M. E. (2003). Subgrouping poor readers on the basis of individual differences in reading-related abilities. *Journal of Learning Disabilities, 36*, 151 – 164.

Chapman, L. J., Chapman, J. P., Curran, T. E., & Miller, M. B. (1994). Do children and the elderly show heightened semantic priming? How to answer the question. *Developmental Review, 14*, 159 – 185.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). A dual-route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*, 204 – 256.

DeFries, J. C., Filipek, P. A., Fulker, D. W., Olson, R. K., Pennington, B. F., & Smith, S. D. (1997). Colorado Learning Disabilities Research Center. *Learning Disabilities, 8*, 7 – 19.

Dunn, L. M., & Markwardt, F. C. (1970). *Examiner's manual: Peabody Individual Achievement Test*. Circle Pines, MN: American Guidance Service.

Forster, K. (1976). Accessing the mental lexicon. In R. J. Wales & E. Walker (Eds.), *New approaches to language mechanisms* (pp. 257 – 287). Amsterdam, Netherlands: North Holland.

Forster, K., & Hector, J. (2002). Cascaded versus noncascaded models of lexical and semantic processing the turple effect. *Memory & Cognition, 30*, 1106 – 1116.

Frost, S. J., Mencl, W. E., Sandak, R., Moore, D. L., Rueckl, J. G., Katz, L., et al. (2005). A functional magnetic resonance imaging study of the tradeoff between semantics and phonology in reading aloud. *NeuroReport, 16*, 621 – 624.

Gonnerman, L. M., Seidenberg, M. S., & Andersen, E. S. (2007). Graded semantic and phonological similarity effects in priming: Evidence for a distributed connectionist approach to morphology. *Journal of Experimental Psychology: General, 136*, 323 – 345.

Keenan, J. M. (2000, May). *Additivity in lexical priming*. Talk presented at the Midwestern Psychological Association Annual Meeting, Chicago.

Keenan, J. M., & Betjemann, R. S. (2001, June). *Combinatorial priming in skilled reading and dyslexia.* Poster presented at the meeting of the Society for the Scientific Study of Reading, Boulder, CO.

Keenan, J. M., & Betjemann, R. S. (2002, May). *Combinatorial priming: assessing the architecture of the lexicon in skilled reading and dyslexia*. Poster presented at annual meeting of the Midwestern Psychological Association, Chicago.

Keenan, J. M., & Betjemann, R. S. (2007). Comprehension of single words: The role of semantics in reading and reading disability. In E. Grigorenko & A. Naples (Eds.), *Single word reading: Behavioral and biological perspectives.* Mahwah, NJ: Erlbaum.

Keenan, J. M., Grubb, J., & Betjemann, R. S. (1999, June). *Combinatorial influences of graphemic, phonological, and semantic priming*. Poster presented at the annual meeting of the American Psychological, Denver, CO.

Keller, R. A. (1981). *Semantic and phonemic/graphemic influences on lexical decisions*. Unpublished doctoral dissertation, University of Denver, Denver, CO.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction integration model. *Psychological Review, 95*, 163 – 182.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge University Press.

Lovrich, D., Cheng, J. C., Velting, D. M., & Kazmerski, V. (1997). Auditory ERPs during rhyme and semantic processing: Effects of reading ability in college students. *Journal of Clinical and Experimental Neuropsychology, 19*, 313–330.

Mann, V. A., Liberman, I. Y., & Shankweiler, D. (1980). Children's memory for sentences and word strings in relation to reading ability. *Memory & Cognition, 8*, 329–335.

Mann, V. A., Shankweiler, D., & Smith, S. T. (1984). The association between comprehension of spoken sentences and early reading ability: The role of phonetic representation. *Journal of Child Language, 11*, 627–643.

Merrill, E. C., Sperber, R. D., & McCauley, C. (1980). The effects of context on word identification in good and poor readers. *Journal of Psychology, 106*, 179–192.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing words: Evidence of a dependence upon retrieval operations. *Journal of Experimental Psychology, 90*, 227–234.

Nation, K., & Snowling, M. J. (1998). Individual differences in contextual facilitation: Evidence from dyslexia and poor reading comprehension. *Child Development, 69*, 996–1011.

Nation, K., & Snowling, M. J. (1999). Developmental differences in sensitivity to semantic relations among good and poor comprehenders: Evidence from semantic priming. *Cognition, 70*, B1–B13.

Olson, R. K. (2004). SSSR, environment, and genes. *Scientific Studies of Reading, 8*, 111–124.

Olson, R. K., Forsberg, H., Wise, B., & Rack, J. (1994). Measurement of word recognition, orthographic, and phonological skills. In G. R. Lyon (Ed.), *Frames of reference for the assessment of learning disabilities: New views on measurement issues* (pp. 243–277). Baltimore: Brookes.

Pennington, B. F. (2006). From single to multiple deficit models of developmental disorders. *Cognition, 101*, 385–413.

Perfetti, C. A. (1985). *Reading ability*. New York: Oxford University Press.

Perfetti, C. A., Marron, M. A., & Foltz, P. W. (1996). Sources of comprehension failure: Theoretical perspectives and case studies. In C. Cornoldi & J. Oakhill (Eds.), *Reading comprehension difficulties: Processes and intervention* (pp. 137–165). Mahwah, NJ: Erlbaum.

Perfetti, C. A., & Roth, S. (1981). Some of the interactive processes in reading and their role in reading skill. In A. Lesgold & C. A. Perfetti (Eds.), *Interactive processes in reading* (pp. 269–297). Hillsdale, NJ: Erlbaum.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103*, 56–115.

Purvis, K. L., & Tannock, R. (1997). Language abilities in children with attention deficit hyperactivity disorder, and normal controls. *Journal of Abnormal Child Psychology, 25*, 133–144.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96*, 523–568.

Seidenberg, M. S., & Tanenhaus, M. K. (1979). Orthographic effects on rhyme monitoring. *Journal of Experimental Psychology: Human Learning & Memory, 5*, 546–554.

Shanahan, M. A., Pennington, B. F., Yerys, B. E., Scott, A., Boada, R., Willcutt, E. G., et al. (2006). Processing speed deficits in attention deficit/hyperactivity disorder and reading disability. *Journal of Abnormal Child Psychology, 34*, 584–601.

Simpson, G. B., & Lorsbach, T. C. (1987). Automatic and conscious context effects in average and advanced readers. *Journal of Research in Reading, 10*, 102–112.

Simpson, G. B., Lorsbach, T., & Whitehouse, D. (1983). Encoding and contextual components of word recognition in good and poor readers. *Journal of Experimental Child Psychology, 35*, 161–171.

Snowling, M. J. (1998). Dyslexia as a phonological deficit: Evidence and implications. *Child Psychology and Psychiatry Review, 3*, 4–11.

Snowling, M. J. (2000). *Dyslexia* (2nd ed.). Oxford, England: Blackwell.

Spring, C., & French, L. (1990). Identifying children with specific reading disabilities from listening and reading discrepancy scores. *Journal of Learning Disabilities, 23*, 53–58.

Stanovich, K. E. (1980). Toward an interactive-compensatory model if individual differences in the development of reading fluency. *Reading Research Quarterly, 16*, 32–71.

Stanovich, K. E. (1984). The interactive-compensatory model of reading: A confluence of developmental, experimental and educational psychology. *Remedial and Special Education, 5*, 11–19.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*, 360–407.

Stanovich, K. E. (2000). *Progress in understanding reading: Scientific foundations and new frontiers*. New York: Guilford Press.

Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning & Verbal Behavior, 18*, 645–659.

Tiu, R. D., Jr., Thompson, L. A., & Lewis, B. A. (2003). The role of IQ in a component model of reading. *Journal of Learning Disabilities, 36*, 424–436.

Tunmer, W. E., & Chapman, J. W. (1998). Language prediction skill, phonological recoding ability and beginning reading. In C. Hulme & R. M. Joshi (Eds.), *Reading and spelling: Development and disorders* (Vol. 7, pp. 33–67). Mahwah, NJ: Lawrence Erlbaum.

Vargo, F. E., Grosser, G. S., & Spafford, C. S. (1995). Digit span and other WISC-R scores in the diagnosis of dyslexia in children. *Perceptual and Motor Skills, 80*, 1219–1229.

Vellutino, F. R., Scanlon, D. M., & Spearing, D. (1995). Semantic and phonological coding in poor and normal readers. *Journal of Experimental Child Psychology, 59*, 76–123.

Vellutino, F. R., Scanlon, D. M., & Tanzman, M. S. (1988). Lexical memory in poor and normal readers: Developmental differences in the use of category cues. *Canadian Journal of Psychology, 42*, 216–241.

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin, 101*, 192–212.

Wechsler, D. (1974). *Examiners' manual: Wechsler Intelligence Scale for Children—Fourth Edition*. San Antonio, TX: Psychological Corporation.

Willcutt, E. G., Pennington, B. F., Olson, R. K., Chhabildas, N., & Hulslander, J. (2005). Neuropsychological analyses of comorbidity between reading disability and attention deficit hyperactivity disorder: In search of the common deficit. *Developmental Neuropsychology, 27*, 35–78.

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin, 131*, 3–29.

Child Development, July/August 2008, Volume 79, Number 4, Pages 1103–1118

Individual Differences in Adolescents' Beliefs About the Legitimacy of Parental Authority and Their Own Obligation to Obey: A Longitudinal Investigation

Nancy Darling
Oberlin College

Patricio Cumsille and M. Loreto Martínez
Pontificia Universidad Católica de Chile

Adolescents' beliefs about the legitimacy of parental authority and obligation to obey were examined in 568 Chilean adolescents (11–14 years old at Wave 1), followed once a year for 4 years. Adolescents' beliefs about parental legitimacy and obligation to obey declined with age. The steepest decline occurred during early adolescence, particularly in the personal domain. Adolescents who were uninvolved in problem behavior and perceived their parents to be supportive or high in monitoring at Wave 1 were more likely to endorse parental legitimacy and obligation to obey over time. There was little evidence that parenting or problem behavior moderated the normative decline in adolescents' beliefs about parental authority. Findings concerning individual differences in adolescents' endorsement of parental authority are highlighted in this study.

When parents set rules that adolescents disagree with, adolescents usually obey (Darling, Cumsille, & Martínez, 2007; Darling, Cumsille, & Peña-Alampay, 2005). Why? Most research on adolescent socialization has focused on what parents do to facilitate positive adolescent development (Parke & Buriel, 1998), with extensive literatures concerning both stylistic differences in parents' characteristics, such as support, and behavioral and psychological control (e.g., Barber, Stolz, & Olsen, 2005; Pettit, Laird, Dodge, Bates, & Criss, 2001), and on specific parenting practices such as monitoring (e.g., Crouter & Head, 2002). Although developmental researchers have long recognized that effective parenting is made easier with a cooperative child (Thomas & Chess, 1968), relatively less attention has been given to the processes underlying differences in adolescents' willingness to be socialized (although see Kerr, Stattin, & Trost, 1999; Moffitt, 1993; Patterson, DeBaryshe, & Ramsey, 1989, for important exceptions). Understanding these processes is particularly important during adolescence, when normative increases in adolescents' desire for psychological and behavioral autonomy are likely to conflict with continued parental efforts to protect, socialize, and regulate their children (Smetana, 2002). This article examines normative change and individual differences in adolescents' beliefs that parents have the right to set rules about particular areas of their lives (the legitimacy of parental authority) and their beliefs that they must obey parental rules they disagree with (obligation to obey). Both characteristics of adolescents and of their parents are examined as predictors of individual differences in adolescents' beliefs and as moderators of normative change.

Parents' knowledge of their adolescent children's lives, adolescents' decisions to share information about their lives, and adolescents' decisions to obey parents when they disagree with them depend in part upon adolescents' beliefs about the legitimacy of parental authority and their own obligation to obey. Legitimacy of parental authority refers to the extent to which parents' assertion of control over an area is believed to be a natural or an appropriate extension of their role as parents (e.g., Smetana, 1988). Obligation to obey is a related construct, referring to the belief that children are obliged to obey parents, even when they and their parents disagree. Darling and Steinberg (1993) hypothesized that one reason that the adolescents of authoritative parents (i.e., those who are warm and high in behavioral control but low in psychological control) evince positive developmental outcomes is that they (a) are more likely to see their parents as legitimate authorities and believe themselves obliged to obey them, (b) are more likely to internalize their values, and (c) are thus both more open to parental socialization and more likely to make autonomous decisions consonant with parental values. Research on the association of legitimacy beliefs and parenting has been consistent with this

---

This study was funded by Grants 1010933, 7010933, and 1070852 from Fondo Nacional de Desarrollo Científico y Tecnológico, Chile.

Correspondence concerning this article should be addressed to Nancy Darling, Department of Psychology, Severance Hall, Oberlin College, Oberlin, OH 44074. Electronic mail may be sent to nancy.darling@oberlin.edu.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7904-0019

prediction. Early adolescents who believe that their parents have legitimate authority to regulate personal issues, such as who their friends are or how they dress, also report that their parents have more knowledge of their lives (Smetana & Daddis, 2002). Adolescents' beliefs about their obligation to disclose information to parents, as well as greater trust, and in the case of personal issues, more parental acceptance, predicts more disclosure and less secrecy (Smetana, Metzger, Gettman, & Campione-Barr, 2006). Parents whose knowledge of their adolescents' lives stems from voluntary disclosure are most likely to have adolescents with lower levels of problem behavior (Crouter, Bumpus, Davis, & McHale, 2005; Kerr & Stattin, 2000). Further, adolescents' decisions to disclose information to parents depends upon both the adolescents' perception that parents are supportive (Crouter et al., 2005; Darling, Cumsille, Caldwell, & Dowdy, 2006) and, in case of disagreement, adolescents' beliefs that they are obliged to obey parents in spite of disagreement (Darling et al., 2006). Adolescents who express general agreement with parents and who feel obliged to obey parents across a range of issues are more likely to obey parents than are their peers, controlling for parental support and monitoring. Examining variability within adolescents across different issues and controlling for both the presence of rules and the parents' reports of rule enforcement, adolescents are most likely to obey parents in situations where they agree with parents, believe that parents have legitimate authority, and believe that they are obliged to obey parents in spite of disagreement (Darling et al., 2007).

Research on adolescents' and parents' beliefs about the legitimacy of parental authority and the obligation to obey emerged from research on social cognition and moral development (e.g., Turiel, 1983) and has primarily focused on (a) differences in legitimacy beliefs depending upon the domain of the issue in question; (b) normative, age-related differences in adolescents' and parents' legitimacy beliefs; and (c) differences in the legitimacy beliefs of adolescents and parents. Research in the United States in samples of African-, Cuban-, Chinese-, European-, Filipino-, and Mexican-American backgrounds (e.g., Darling et al., 2005; Fuligni, 1998; Smetana, 2000) and in samples from Brazil, Chile, China, Japan, and the Philippines (e.g., Darling et al., 2005; Hasebe, Nucci, & Nucci, 2004; Nucci, Camino, & Sapiro, 1996; Yau & Smetana, 2003) has yielded strikingly consistent results. Adolescents and their parents are more likely to judge issues that are defined as moral (e.g., hitting or stealing), prudential (e.g., drinking, using drugs), and conventional (e.g., talking back to parents, cursing) as legitimately within the domain of parental authority, and to judge issues defined as personal (e.g., choice of friends or extracurricular activities) as outside legitimate parental control (see Smetana, 1997, for a detailed review of this perspective). As adolescents become older, both they and their parents define more issues as falling outside the legitimate domain of parental authority, although parents lag behind adolescents in the speed of this transition (Smetana, 2000). Much of this normative contraction in the legitimate domain of parental authority occurs in the personal domain and in multidomain issues that have personal aspects.

Differences between adolescents' and parents' beliefs about the legitimacy of parental authority have been hypothesized to cause conflict within the family, especially in early and middle adolescence (Smetana, 1996). Parental assertion of control over personal issues is also associated with adolescents' perceptions that parents are exerting inappropriate psychological control (Smetana & Daddis, 2002) and with more internalizing symptoms (Hasebe et al., 2004). Nucci and Smetana have argued that adolescents' assertion of control over personal issues is a fundamental part of the development of healthy autonomy (Nucci, Killen, & Smetana, 1996; Smetana, Crean, & Campione-Barr, 2005). Moderate amounts of resistance to parental authority, or even subversion, may be a sign of healthy development (Smetana et al., 2005). However, premature assertion of personal control (or, conversely, premature rejection of parental legitimacy) over areas normatively defined as legitimately within parental control may be both symptomatic of poor parent–adolescent relationships (Darling et al., 2005) and indicative of problematic developmental pathways (Steinberg & Silverberg, 1986).

Although there is a clear, age-related decline in the range of issues that adolescents and parents believe to be within the legitimate sphere of parental authority, there is considerable within-age variability in adolescents' endorsement of parental authority (Darling et al., 2005). In a recent article in which latent class analyses were used to examine the patterning of adolescents' legitimacy beliefs (Cumsille, Darling, Flaherty, & Martínez, 2006), three distinct patterns emerged. In the first pattern (parent centered), adolescents were likely to grant parents legitimacy over both multidomain and prudential issues and were more likely than their peers to grant parents authority over personal issues. In the second pattern (adolescent centered), adolescents were unlikely to grant parents legitimacy over personal, prudential, or multidomain issues. In the third pattern (shared), adolescents clearly differentiated parental legitimacy across domains and were likely to grant parents authority over prudential issues but unlikely to grant them authority over personal issues. Analogous patterns emerged among early, middle, and late adolescents, with the parent-centered pattern most common at each age (ranging between 49% and 58% of each age group). Although cross-sectional, the proportion of adolescents evincing each pattern was relatively stable across age groups, suggesting that youth showing different patterns may be following different trajectories in the development of legitimacy beliefs.

Individual differences in adolescents' beliefs are associated with differences in adolescents' behavior in ways that may undermine or facilitate parenting. For example, adolescents who are heavy drug users tend to see substance use as a personal rather than prudential issue and thus outside the legitimate sphere of parental authority (Nucci, Guerra, & Lee, 1991). In case of disagreement, adolescents who believe that parents have legitimate authority with regard to specific issues and feel obliged to obey are (a) more likely to disclose information to parents about that issue rather than lie (Darling et al., 2006), (b) more likely to argue about the issue with parents (Cumsille, Darling, & Peña-Alampay, 2002), and (c) more likely to obey (Darling et al., 2007). Further, adolescents who believe that they are obliged to disclose to parents both disclose more to parents and are less likely to keep secrets from them (Smetana et al., 2006). We know little about where these differences in adolescents' beliefs about parental legitimacy and obligation to obey come from or how they may unfold over time. In addition, research on normative changes in adolescents' beliefs has focused on adolescents' beliefs in the legitimacy of parental authority. Much less is known about children's beliefs about their obligation to obey parents when they disagree with them. These beliefs may be particularly important in understanding decision making in situations where direct parental supervision or distal monitoring is difficult or dependent on children's willingness to disclose.

The goals of the present article were (a) to examine longitudinal change in Chilean adolescents' beliefs about the legitimacy of parental authority and their own obligation to obey parents in case of disagreement and (b) to predict individual differences in adolescents' beliefs and changes in those beliefs as a function of their own and their parents' characteristics. Adolescents were in early to middle adolescence at Wave 1 of the study (11–14 years old) and were followed once a year for 4 years. Five specific issues were addressed:

1. Do adolescents become less likely to endorse the legitimacy of parental authority and their own obligation to obey with age?
2. Is the decline in adolescents' endorsement of parental legitimacy and their own obligation to obey uniform across adolescence, or does it decline more rapidly in early adolescence than in middle or later adolescence?
3. Is the decline in endorsement of parental legitimacy and obligation to obey parallel in all domains, or does it decline more quickly in some domains than in others?
4. Are adolescents who perceive their parents to be supportive and high in monitoring more likely to endorse the legitimacy of parental authority and their own obligation to obey than those whose parents do not? Are adolescents who were relatively more involved in problem behavior less likely to endorse parental authority and their own obligation to obey?
5. Do adolescents' beliefs in the legitimacy of parental authority and their own obligation to obey decline less rapidly when they see their parents as supportive and high in monitoring? Do the beliefs of adolescents who are relatively more involved in problem behavior decline more rapidly?

Questions 1–3 provide a partial replication and extension of longitudinal research by Smetana (2000), documenting changes from ages 13 to 15 in a sample of 82 African American adolescents' beliefs about parental authority using a sample of 568 Chilean adolescents. Because adolescents ranged in age from early to middle adolescence at Wave 1 of this study, differences in belief trajectories from early to middle and from middle to late adolescence could be compared. Based on prior research, it was hypothesized that normative declines would be most rapid in the personal and multidimensional domains. Because previous research has suggested that a normative shift in parental authority occurs during the pubertal transition, it was further hypothesized that the decline in endorsement of parental authority would be more marked in the transition from early to middle adolescence than during later periods (Smetana, 1996).

Questions 4 and 5 focus on individual differences and changes in adolescents' beliefs about the legitimacy of parental authority and their own obligation to obey as a function of initial parental support and monitoring and initial adolescent problem behavior. The former question focuses on overall differences in adolescents' beliefs across ages (Which adolescents are most likely to believe that parents have legitimate authority and that they are obliged to obey?).

Although research has consistently shown that monitoring and support are associated with more positive development outcomes (Steinberg & Silk, 2002), the underlying processes linking parent attributes and adolescent outcomes have been studied less frequently, and until recently, the empirical literature has neglected the adolescent as an active participant in the processes. It was hypothesized that adolescents who see their parents as supportive and expressing interest in their adolescents' lives through monitoring will be more likely to believe that parents have legitimate authority and themselves as obliged to obey (Darling & Steinberg, 1993) and that adolescents who are relatively more involved in problem behavior will be less likely to do so (Jessor & Jessor, 1977; Kerr et al., 1999; Moffitt, 1993; Nucci et al., 1991).

Question 5 focuses on individual differences in the trajectories of adolescents' beliefs over time. It is possible that children come into adolescence with different beliefs about parental authority and that these relative differences remain stable over time. For example, adolescents who see their parents as supportive may also see them as more legitimate, but the decline in legitimacy endorsement over time may be parallel for adolescents with supportive and less supportive parents. Alternatively, seeing parents as supportive may slow the rate at which adolescents' belief in the legitimacy of parental authority declines. These two models of change have very different implications for understanding the association of parenting and adolescents' characteristics with individual differences in adolescents' legitimacy beliefs.

The Chilean Family Context

Chilean adolescents' relations with their parents have not been studied extensively. Chilean culture has been changing rapidly due to urbanization and outside-media influences from Europe, the United States, and other Latin American countries (Instituto Nacional de la Juventud, 1999). In addition, Chile is in the midst of a profound cultural shift resulting from the fall of the military dictatorship and the reinstitution of democracy in 1990. Political changes, combined with rapid economic development, have resulted in a sharp rejection of authoritarian values and liberalization of social and sexual norms (Martínez, Cumsille, & Thibaut, 2006). Reflecting these changes, contemporary Chilean parents appear to be relatively uncomfortable with direct assertion of control and rule enforcement and report that they are less authoritarian and less likely to use power-assertive parenting techniques than did their own parents (Martínez et al., 2006). For example, parents reported discussing areas of parent–adolescent disagreement but leaving the decision to adolescents 78% of the time, letting the adolescent do as they wanted 11% of the time, and enforcing rules only 11% of the time (Darling et al., 2007). Chilean youth appear to share this discomfort with parental control. Chilean youth are less likely than Filipino or U.S. youth to (a) believe that their parents have the right to govern their behavior (particularly with regard to prudential areas, such as alcohol use or curfew), (b) believe that they are obliged to obey parents when they disagree, or (c) agree with parents (Darling et al., 2005). More than 85% of Chilean youth think of themselves as different from the adult generation both in how they think and in how they act (Martínez et al., 2006). Across all social classes, younger Chileans espouse more individualistic values than older cohorts, although the society remains highly family oriented (Programa de Naciones Unidas para el Desarollo, 2002) Taken together, these findings suggest that family relations in contemporary Chile are typically close but that both parents and adolescents tend to expect less direct assertion of parental control and less strict adolescent obedience than might be typical in the contemporary United States.

Although Chile remains a socially conservative country in many ways (e.g., divorce was only legalized in the country in 2005), political and social changes in Chile have been accompanied by problems usually associated with urban youth in Western countries, including relatively high rates of alcohol use (the legal drinking age is 18), substance use, vandalism, and petty theft (Darling et al., 2007; Martínez et al., 2006). In a nationally representative sample, 83% of Chilean youth reported onset of sexual intercourse prior to age 20, 15% reporting onset prior to age 15, and 24% of 15- to 19-year-olds reported that their last sexual encounter was in a casual relationship rather than with a boyfriend, girlfriend, or fiancé (Instituto Nacional de la Juventud, 1999). Despite these changes, adolescents tend to hold fairly traditional ideas about gender roles within the family. These beliefs about the division of parental responsibilities may reflect the fact that fewer Chilean women participate in the paid labor force than in any other South American country (39%; Martínez et al., 2006).

Method

Participants and Procedure

The sample included 568 sixth to ninth graders recruited to participate in a longitudinal study of parenting, adolescent risk, and resilience (mean Wave 1 age = 12.95 years, *SD* = 0.99, range = 11–14 years), each of whom had participated in at least two waves of the 4-year study. Participants were recruited from seven schools from two counties in Santiago, Chile (92% of 14- to 17-year-old Chileans are enrolled in school; Martínez et al., 2006). Schools are typically organized from Grades 1 to 8 and from Grades 9 to 12; therefore, sixth and seventh graders experienced a transition from primary to secondary school during the study's duration. Schools were selected to represent the diversity of socioeconomic status (SES) observed in the Chilean educational system and included two private, one subsidized private, and four public schools, thus oversampling for lower SES students. Fifty-six percent of the sample were female. Seventy-four percent of the adolescents lived with both biological parents at the time of the first assessment.

Public schools were contacted through the board of education of each county. Private schools were contacted directly. Consistent with Chilean regulations and research practices, active adolescent assent and passive parental consent were used. After mailing descriptions of the study home to parents, students were given oral and written information about the study's purpose and procedure, invited to sign written consent forms, and completed questionnaires within their classrooms. Consent procedures and survey administration were carried out by upper-level psychology students trained by the researchers. Participants received a candy bar as a thank you for their participation. During each wave of the study, fewer than 2% of students declined to participate or had permission withheld by parents. Subsequent survey administrations took place approximately 1 year apart. Because of budgetary limitations, follow-up data collection for each wave was limited to only a randomly selected subset of the original sample. During follow-up, classrooms within schools and then individuals within classrooms were selected at random. In addition, because student contact and data collection were entirely school based, somewhat less than 5% of total data loss occurred due to absences or because students changed or left school during the course of the study. For these analyses, data are included for all individuals who participated in at least two waves of data collection, including both Wave 1 and Wave 2, when predictor variables (parental monitoring and support and substance use) were measured. Thirty-three percent of students provided data for 2 years of the study, 44% for 3 years, and 23% for 4 years. Years of participation did not vary by age, family type, adolescent problem behavior, or parental support or monitoring ($p > .05$).

Adolescents provided information on demographic background, attitudes toward school and civic involvement, activities, self-esteem, problem behavior, parenting practices, and risk and protective factors. In addition, adolescents each completed a Strategic Disclosure Questionnaire (Darling et al., 2007), with adolescents answering six questions about each of 20 different issues. Issues were selected to represent areas that are frequent sources of parent-adolescent disagreement and were identified through review of the extant literature and focus groups. Questionnaires were developed jointly by native American English and Chilean speakers, with the translation of individual items discussed and reviewed by bilingual, native Chilean speakers. Cultural appropriateness of stems and questions were validated in a separate pilot study involving more than 200 Chilean adolescents. Measures are described here in English.

Measures

Legitimacy of parental authority and obligation to obey. To assess adolescents' legitimacy beliefs, adolescents were asked, "Is it okay for your parents to set rules about this issue?" for each of the 20 issues on the Strategic Disclosure Questionnaire (0 = *no*, 1 = *yes*). Obligation to obey was assessed by asking adolescents "If you disagree with your parents, do you have to obey?" (0 = *no*, 1 = *yes*). Across issues and wave of data collection, adolescents reported that they believed that parents had the right to set rules for 57.7% of issues and that they were obliged to obey for 59.9% of issues. As expected, adolescents' beliefs about legitimacy and obligation to obey were related, although not identical. If an adolescent endorsed parental legitimacy for a given issue, they also endorsed their own obligation to obey 83.7% of the time. If adolescents did not endorse legitimacy for an issue, they reported that they were obliged to obey only 27.7% of the time.

Between-Person Predictors

Parental monitoring. At Wave 1, adolescents reported on the extent to which their mothers and fathers tried to know (a) who their friends were, (b) what they did after school, (c) what they did in the evenings, (d) what they did with their free time, (e) how they spent their money, and (f) how they were doing in school. Adolescents responded on a 1–3 scale, with 3 indicating *parents tried a lot* and 1 indicating *parents did not try.* A mean score was calculated (mother $\alpha = .80$, father $\alpha = .86$), and the maximum of either mother or father monitoring was used as an indicator of parental monitoring ($M = 2.47$, $SD = 0.49$). The decision to use the maximum of mother or father score rather than a mean was made because the normative difference in ratings of maternal and paternal monitoring and the amount of missing father data resulted in an artificial inflation of monitoring scores for those who reported on only one parent.

Parental support. Because parental support was not measured at Wave 1, a Wave 2 assessment was used. Using a 5-point scale, adolescents reported on the extent to which they agreed with four items describing each of their parents (sample items: "My mother/father likes to spend time just talking with me" and "I can count on my mother/father to help me with my problems"; mother $\alpha = .80$, father $\alpha = .83$). Items were recoded such that high scores indicated high support. Mean scores were calculated for each parent. The maximum of mother or father score was used as an indicator of parental support ($M = 4.36$, $SD = 0.79$) to reduce artificial bias favoring adolescents who reported on only one parent.

Problem behavior. At Wave 1, adolescents reported on how frequently they had engaged in five problem behaviors from *never in my life* (1) to *more than once in the last month* (5). Sample items included damaged public or private property, taken something from a store without paying for it, been in a fist fight, brought a weapon to school, or been in a fight that resulted in an injury in which someone needed to be seen by a doctor or nurse ($M = 1.31$, $SD = 0.59$). (Substance use was excluded because of normative age differences in use.)

Within-Person Predictors

Time. Time was coded from 0 to 3, with 0 indicating *baseline assessment* and 3 indicating *Wave 4 of the study.*

Domain. Because issues in this study were chosen to reflect common sources of parent–adolescent disagreement, rather than a specific theoretical model, issue domain membership was empirically derived from factor analysis of adolescents' ratings of parental legitimacy using a separate sample (Darling et al., 2005). Each issue was classified into one of five domains: personal (choice of friends, what you wear, use of free time, extracurricular activities, use of money), prudential (smoking, alcohol, and drug use), parent expectations (time on the telephone, homework, school performance), opposite-sex relations ("Your relationships with the opposite sex [phone calls, going to dances or out with mixed sex groups, dating]" and "Your relationship with your boyfriend or girlfriend [time you spend together, privacy you are allowed, how serious you are, your sexual relationship]"), and multidomain (type of video viewed, spending time with people parents do not like, unsupervised time with friends, where you go with friends, hanging out after school and after dinner, curfew). Empirical classification differs from prior theoretical research on domains primarily in the separation of the opposite-sex relations and personal domains. Issues classified here in the opposite-sex relations domain are quite similar to those classified as multifaceted in stimulus material currently used by Smetana (e.g., Smetana et al., 2006).

Analytic Strategy

A series of analyses was undertaken predicting (a) change in adolescents' beliefs as a function of age and time, (b) change in adolescents' beliefs as a function of age and time run separately by domain, and (c) individual differences in adolescents' beliefs as a function of age, time, parenting, and adolescent problem behavior. Analysis of these data was complicated by the nonindependence of each participant's assessment of 20 issues for up to four waves of data collection, which violates the assumptions of techniques such as multiple regression, and thus artificially inflates error terms. In this study, data were available from 568 adolescents reporting on 22,589 issues over the four waves. Hierarchical general linear modeling (HGLM) was specifically designed to adjust the degrees of freedom in the model to compensate for nonindependence of observations (Raudenbush & Bryk, 2002). In these analyses, HGLM was used to parse variance in adolescents' beliefs into a participant component (differences between different adolescents' beliefs predicted from adolescents' age and gender, parental monitoring, parental support, and adolescent problem behaviors) and a within-adolescent issues component (differences in beliefs with regard to specific issues predicted from domain membership and time). Because of its structure, all within-person HGLM analyses effectively control for between-person differences. In these analyses, therefore, analyses predicting within-person differences in legitimacy beliefs and obligation to obey from time and domain controlled for between-person differences in gender, age, and initial perceived parental support and monitoring and adolescent problem behavior. All HGLM analyses were performed using the Bernoulli procedure with logit link for dichotomous outcomes, such that predictors discriminate between two dichotomous choices. Because Level 1 variance is heterogeneous in Bernoulli logit-link analyses (Snijders & Bosker, 1999), the percentage of variance attributed to each level and explained by each model was not calculated. Unit-specific models are reported.

Results

Legitimacy of Parental Authority and Obligation to Obey as a Function of Age, Time, and Domain

A series of two-level (issue within adolescent) HGLM Bernoulli logit-link analyses were used to test the hypotheses that (a) older adolescents would be less likely to endorse the legitimacy of parental authority and (b) belief in the legitimacy of parental authority would decrease over time. In these analyses, Wave 1 age was used to predict between-adolescent differences in adolescents' beliefs about legitimacy. Time (Years 0, 1, 2, 3) was used to predict within-adolescent differences in beliefs. The interaction between Wave 1 age and time tested the hypothesis that the time-related decline in likelihood is steeper for younger adolescents than for older adolescents. Analyses were first run for all issues and then run separately by domain. A series of three-level HGLM models (issue within domain within person) were used to test for domain difference in the intercepts, age, and time model coefficients (Level 1 $df = 22589$, Level 2 $df = 2752$, Level 3 $df = 550$). In preliminary analyses, a curvilinear time function (time^2) was also added to the models, with mixed results. Based on the recommendations of Raudenbush and Bryk (2002) regarding the fitting of longitudinal models with a small number of measurement points and the significant interaction of initial age and time of measurement, the decision was made to use the more parsimonious linear model for the final analyses. Curvilinear results are very similar to those presented here and are available from the first author.

The predicted probabilities of endorsing parental legitimacy and obligation to obey for all items are graphed in Figure 1. In this figure, separate cohorts are indicated by different lines (i.e., adolescents in the study from ages 11 to 14, 12 to 15, 13 to 16, and 14 to 17). The likelihood of endorsing parental legitimacy and the obligation to obey decreased with both initial age and time ($ps \leq .000$). Adolescents who entered the study at an older age were less likely to endorse parental legitimacy or their own obligation to obey. Predicting within-person change, endorsement decreased with time. The interactions between both legitimacy and obligation to obey with Wave 1 age were significant ($ps \leq .000$), such that the within-person decrease in legitimacy over time was greater for those adolescents who had started the study at a younger age.

Results of both the two- and three-level HGLM models calculated separately by domain are reported in Table 1. The likelihood that adolescents would endorse parental authority varied by domain ($p \leq .01$), with adolescents most likely to endorse parental authority in the prudential domain and least likely to endorse parental authority in the personal domain. In all domains, adolescents who were youngest at Wave 1 were most likely to endorse parental authority ($p \leq .05$). Within adolescents, the likelihood of endorsing parental authority declined with time ($p \leq .05$). The difference in endorsement of legitimacy between older and younger adolescents was largest in the prudential domain and smallest for multidomain issues. The within-adolescent decline in legitimacy with time was also steepest in the prudential domain and shallowest for multidomain issues. In the personal domain only, the decline in endorsement of legitimacy was steeper for younger adolescents than for older adolescents ($p \leq .01$).

Identical analyses were carried out predicting between- and within-adolescent differences in the likelihood that adolescents believed that they were obliged to obey parents in case of disagreement. Results of the analyses predicting obligation to obey paralleled those predicting legitimacy of parental

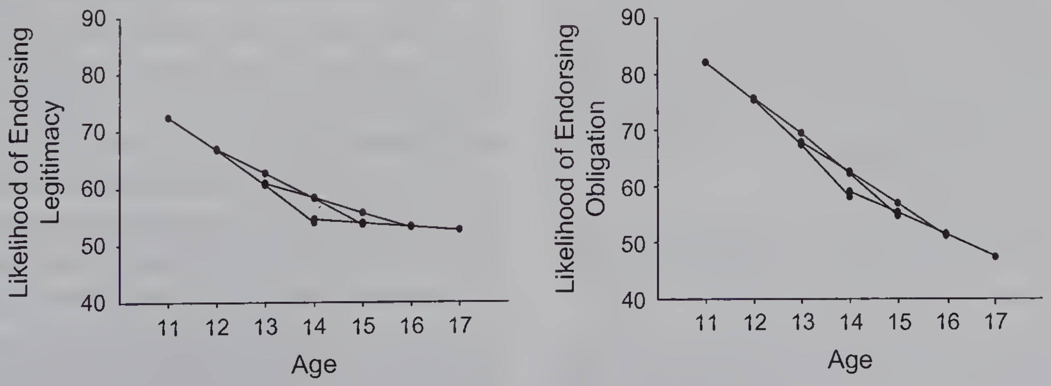


Figure 1. Predicted likelihood of endorsing parental legitimacy and obligation to obey by age.
Note. Separate lines indicate adolescents who began the study at different ages.

Table 1

Results of HGLM Analyses Predicting Adolescents' Beliefs That Issues Are Within the Legitimate Domain of Parental Authority and That They Are Obliged to Obey From Age, Time, and Domain

	Domain				
	Personal	Opposite-sex relations	Prudential	Expectations	Multidomain
			Legitimacy		
Between person					
Intercept	-0.42^{***}_{a}	0.24^{**}_{b}	2.06^{***}_{c}	1.26^{***}_{d}	0.59^{***}_{e}
Wave 1 Age	-0.32^{***}_{ac}	-0.21^{*}_{ab}	-0.43^{***}_{c}	-0.33^{***}_{ac}	-0.22^{***}_{b}
Within person					
Time	-0.16^{***}_{ac}	-0.19^{***}_{ac}	-0.27^{***}_{c}	-0.11^{*}_{ab}	-0.06^{*}_{b}
Time × Wave 1 Age	0.09^{**}	0.09	0.10	0.09	0.10
			Obligation to obey		
Between person					
Intercept	0.15^{*}_{a}	0.49^{***}_{b}	1.93^{***}_{c}	1.35^{***}_{d}	0.89^{***}_{e}
Wave 1 Age	-0.51^{***}_{ac}	-0.34^{***}_{b}	-0.54^{***}_{a}	-0.19^{*}_{bc}	-0.32^{***}_{b}
Within person					
Time	-0.37^{***}_{a}	-0.26^{***}_{bd}	-0.40^{***}_{ab}	-0.17^{***}_{cd}	-0.15^{***}_{c}
Time × Wave 1 Age	0.13^{***}	0.08	0.05	−0.06	0.08^{**}
			Two-level model Approximate *df*		
Between person	559	558	557	559	559
Within person	7217	2844	2844	2887	7175

Note. Positive coefficients indicate greater likelihood of endorsing legitimacy and obligation to obey. Results of the three-level HGLM analyses comparing the coefficients of models of different domains are summarized in the subscript. Within rows, coefficients that share lowercase subscripts are not significantly different from one another ($p > .05$). HGLM = hierarchical general linear modeling.
$^{*}p \leq .05$. $^{**}p \leq .01$. $^{***}p \leq .001$.

authority (Table 1, bottom). Adolescents were most likely to endorse their own obligation to obey for issues in the prudential domain and least likely to endorse these beliefs in the personal domain ($p \leq .001$). In all domains, younger adolescents were more likely to believe that they were obliged to obey parents than were older adolescents ($p \leq .05$), although the age-related differences varied by domain. Interestingly, although adolescents were least likely to feel obliged to obey in the personal domain, this domain also showed a relatively large age-related difference in endorsement of parental legitimacy. Predicting within-adolescent change, the likelihood that adolescents believed that they were obliged to obey declined with time in all domains ($p \leq .001$). For personal and multidomain issues, endorsement of obligation to obey declined more steeply for younger adolescents than for older adolescents ($p \leq .01$).

Together, these analyses confirm previously reported differences in adolescents' beliefs about the legitimacy of parental authority and their own obligation to obey. Cross-sectional age differences were replicated by within-person declines in endorsements with time. Overall, the decline in the likelihood that adolescents would endorse the legitimacy of parental authority and their own obligation to obey was steeper for younger adolescents than for older adolescents. However, this interaction was only statistically significant in the personal domain and, for obligation to obey, for multidomain issues.

Individual Differences in Adolescents' Beliefs as a Function of Parents' and Adolescents' Characteristics

A series of two-level (issue within adolescent) HGLM Bernoulli logit-link analyses were conducted to examine the hypotheses that (a) adolescents who initially described their parents as relatively supportive and high in monitoring would be more likely to grant parents legitimacy and feel that they were obliged to obey over the course of the study and (b) adolescents who were involved in relatively high initial rates of problem behaviors would be less likely to do so. As in the previous analyses, analyses were first conducted for all items and then run separately by domain. Between-adolescent differences in adolescents' beliefs were predicted from parental support, parental monitoring, and problem behavior, controlling for initial age and gender. As in the previous analyses, within-adolescent differences were predicted by time. We also tested hypotheses that the decline in endorsement over time would be smaller for older adolescents and for adolescents whose parents were relatively supportive and high in monitoring, and steeper for adolescents who were initially more involved in problem behavior. In these analyses, variables predicting between-adolescent differences in legitimacy beliefs were centered around the grand mean. Time was entered uncentered, with 0 representing the first wave of data collection. Because previous research indicated that beliefs about parental legitimacy varied by parents' marital status and SES (i.e., Nucci, Camino, et al., 1996; Smetana, 2000; Smetana, Yau, Restrepo, & Braeges, 1991), preliminary analyses were undertaken in which parents' marital status and education were used as predictors. The variables neither predicted adolescents' beliefs nor moderated the association of age or time with beliefs in this sample, and these variables were dropped from subsequent analyses.

Results of the combined model are illustrated in Figure 2. Differences in the predicted likelihood of endorsing legitimacy and obligation are plotted for adolescents who were high in problem behavior and whose parents were low in perceived warmth and monitoring (high risk) and for a comparison group of adolescents who were low in problem behavior and whose parents were high in perceived warmth and monitoring (low risk). High and low problem behavior, warmth, and monitoring were defined by the 25th and 75th percentile. In this figure, separate cohorts are indicated by different lines. In the combined model, younger adolescents were more likely to grant parents legitimate authority than were older adolescents and, looking within adolescents, grew less likely to do so with time ($ps \leq .001$). As in the previous analyses, the decline in endorsement of parental authority was steeper for younger adolescents than for older adolescents ($ps \leq .001$). Both higher initial monitoring and support were associated with greater likelihood of endorsing parental legitimacy ($p \leq .005$) and obligation to obey ($p \leq .05$) over the course of the study. Adolescents who initially reported higher rates of problem behavior were less likely to endorse parental legitimacy or obligation to obey ($p \leq .001$). The decline in endorsement with time did not vary by initial perceived monitoring or warmth ($p > .09$). The decline in adolescents' endorsement of the legitimacy of parental authority (but not obligation to obey) was less steep for those initially high in problem behavior ($p = .005$).

Results of the two-level HGLM models calculated separately by domain are reported in Table 2. In all domains, younger adolescents were more likely to grant parents legitimate authority than were older adolescents and, looking within adolescents, grew less likely to do so with time ($p \leq .01$). For personal and multidomain issues, the decline in likelihood of endorsement with time was steeper among the younger adolescents than among the older adolescents ($p \leq .01$). A similar trend was found for opposite-sex relations issues ($p = .055$). Boys were more likely to grant parents authority over issues in the personal and expectations domains but less likely than girls to grant parents authority over issues in the opposite-sex relations domain.

Although parenting predicted legitimacy beliefs in all domains except for opposite-sex relations, there was little evidence that perceived parents' characteristics moderated the decline in adolescents' legitimacy beliefs with time. Adolescents who described their parents as relatively supportive were more likely than their peers to believe that parents had legitimate authority over personal, prudential, and multidomain issues ($p \leq .05$). Adolescents who described their parents as relatively high in monitoring were more likely to believe that parents had legitimate authority over issues in the

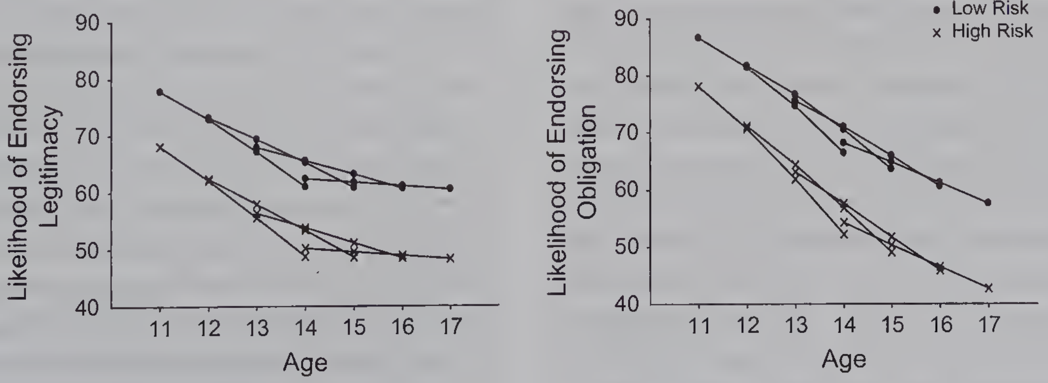


Figure 2. Predicted likelihood of endorsing parental legitimacy and obligation to obey by age. Trajectories of adolescents who are low in parental monitoring and warmth and high in problem behavior (high risk) and adolescents who are high in monitoring and warmth and low in problem behavior (low risk).

Note. Separate lines indicate adolescents who began the study at different ages.

Table 2

Results of HGLM Analyses Predicting Adolescents' Beliefs That Issues Are Within the Legitimate Domain of Parental Authority and That They Are Obligated to Obey Parents From Sex, and Wave 1 Age, Parental Monitoring, Parental Support, and Adolescent Problem Behavior

	Domain				
	Personal	Opposite-sex relations	Prudential	Expectations	Multidomain
		Legitimacy of parental authority			
Between person					
Intercept	−0.42	0.25	2.10	1.29	0.62
Sex	0.23*	−0.70**	0.24	0.37**	−0.18
Wave 1 Age	−0.30**	−0.21**	−0.42**	−0.33**	−0.22**
Parental Monitoring	0.25	0.15	0.16	0.48**	0.36**
Support	0.18*	0.18	0.24*	0.10	0.30**
Problem Behavior	−0.37**	−0.34*	−0.47**	−0.17	−0.47**
Within person					
Time	−0.16**	−0.19**	−0.28**	−0.12**	−0.07**
Time × Wave 1 Age	0.09**	0.09	0.10	0.09	0.10**
Time × Monitoring	0.03	0.03	0.06	−0.27**	−0.06
Time × Support	0.00	−0.05	−0.01	0.04	−0.06
Time × Problem Behavior	0.02	0.03	0.04	0.14	0.15**
		Obligation to obey			
Between person					
Intercept	0.15	0.52	1.99	1.37	0.90
Sex	0.20	−0.68***	0.33*	0.41**	−0.25*
Wave 1 Age	−0.49***	−0.35***	−0.54***	−0.18*	−0.32***
Parental Monitoring	0.22	0.25	0.45*	0.41*	0.29*
Support	0.13	0.26*	0.20	0.07	0.16*
Problem Behavior	−0.45***	−0.63***	−0.83***	−0.33*	−0.61***
Within person					
Time	−0.37***	−0.28*	−0.42*	−0.19***	−0.16***
Time × Wave 1 Age	0.13***	0.09	0.06	−0.06	0.09**
Time × Monitoring	−0.02	−0.07	−0.08	−0.19	−0.06
Time × Support	−0.02	−0.06	−0.06	0.07	−0.01
Time × Problem Behavior	−0.03	0.02	0.06	0.10	0.02

Note. Positive coefficients indicate greater likelihood of endorsing legitimacy. HGLM = hierarchical general linear modeling.
$^{*}p \le .05$. $^{**}p \le .01$. $^{***}p \le .001$.

expectations domain and over multidomain issues ($p \le .01$). Of the 10 interactions between parenting and time, only 1 was statistically significant. In the parent expectations domain only, the decline in the likelihood of endorsing parental legitimacy with time was steeper for adolescents who described their parents as high in monitoring than for their peers ($p \le .01$). As hypothesized, adolescents who were initially more involved in problem behavior were less likely to endorse parental legitimacy for personal, opposite-sex relations, prudential, and multidomain issues ($p \le .05$). For multidomain issues only, adolescents who were involved in problem behavior were less likely to endorse parental authority, but their beliefs declined less over time as well ($p \le .01$), indicating a potential floor effect.

Results of the domain-specific analyses predicting adolescents' obligation to obey are reported in the bottom of Table 2. In all domains, younger adolescents were more likely to believe that they were obliged to obey parents in case of disagreement than were their older peers ($p \le .001$). Within persons, adolescents grew less likely to feel obliged to obey with time ($p \le .001$). For both personal and multidomain issues, the decline over time in likelihood of believing one was obliged to obey was steeper for younger adolescents than for older adolescents ($p \le .001$). As hypothesized, in the opposite-sex relations, prudential, and expectations domains and for multidomain issues, adolescents who reported positive parenting were relatively more likely to feel obliged to obey than their peers. Greater likelihood of endorsing their own obligation to obey was predicted by higher parental monitoring for prudential, expectations, and multidomain issues ($p \leq .05$). Higher parental support was associated with greater likelihood of believing one was obliged to obey for opposite-sex relations and multidomain issues ($p \leq .05$). Adolescents' beliefs about personal issues were not predicted by perceived parenting. Contrary to our hypothesis, parenting did not moderate changes in adolescents' beliefs about their own obligation to obey over time. As hypothesized, adolescents who were initially more involved in problem behavior were less likely to believe that they were obliged to obey parents in case of disagreement ($p \leq .05$). Contrary to our hypothesis, when examined by domain, initial problem behavior did not moderate change in beliefs over time.

Discussion

The purpose of this article was to examine normative change and individual differences in adolescents' beliefs about the legitimacy of parental authority and obligation to obey. These findings replicate and confirm prior cross-sectional (e.g., Darling et al., 2005; Nucci, Camino et al., 1996; Smetana, 1988; Yau & Smetana, 2003) and longitudinal (Smetana, 2000) research showing that as adolescents grow older, they are less likely to believe that issues are within the legitimate domain of parental authority and that they are obliged to obey parents in case of disagreement. This consistency with prior research reinforces interpretation of the contraction of adolescents' understandings of the legitimate domain of parental authority as a reflection of the development of adolescent autonomy, effectance, and agency that is consistent across cultures (e.g., Nucci, Killen, et al., 1996; Smetana et al., 2005). It is notable that these findings have now been replicated in cultures that are strict and parent centered (e.g., China; Yau & Smetana, 2003) as well as those that are relatively low in adolescents' endorsement of parental legitimacy and obligation to obey and in which parents are reluctant to exert control (e.g., Chile; Darling et al., 2005, 2007).

Overall, the sharpest decline in adolescents' endorsement of parental legitimacy and obligation to obey occurs during early adolescence in our sample, growing less steep as adolescents move from middle into later adolescence. Early adolescence is the period during which the gap between adolescents' and their parents' beliefs about legitimacy is largest (Smetana, Yau, & Hanson, 1991). Rapid change in adolescents' beliefs during this period may be one of many factors contributing to the realignment of power relations during the pubertal transition (Collins, 1992). Although adolescents' endorsement of parental legitimacy and their own obligation to obey declines for issues in all domains, the relatively steep decline during early adolescence is seen most clearly in the personal domain for parental legitimacy and for personal and multidomain issues with regard to obligation to obey. Six of the seven multidomain issues combine elements of the personal and prudential domains (e.g., "spending time with people your parents don't like" and "spending time with friends when no adults are around"). Prior work has shown that early adolescents broaden the number of issues they consider to be within the personal domain and that this change is largely responsible for the difference between the boundaries adolescents and parents set for legitimate authority (e.g., Smetana, Daddis, & Chuang, 2003). In particular, adolescents reason about issues that are conceptually defined as multidomain as if they were personal.

In addition to the biological and cognitive changes associated with the transition from early to middle adolescence, it is important to note that social transitions are also likely to contribute to changes in beliefs during this period. In Chile, the transition from early to middle adolescence is marked by a transition from primary to secondary school. Chilean youth also begin spending considerable time socializing with friends on weekends during early adolescence, particularly during the late evening and early morning hours. Although most socializing occurs in family homes (Martínez et al., 2006), adolescent gatherings become both less family centered and less supervised at this age. Spending times in these settings may thus provide an impetus for changing adolescents' beliefs about parental legitimacy and their own obligation to obey, particularly with regard to issues that have elements that are both personal and prudential. More detailed longitudinal analyses that focus on specific contextual and behavioral changes and their temporal relationship to changing legitimacy beliefs would shed additional light on the processes underlying normative change.

The present article extends previous research by focusing on how characteristics of both parents and adolescents predict individual differences in the trajectory of normative change in adolescents' beliefs over time. Adolescents who perceived their parents to be supportive and high in monitoring at the beginning of the study were more likely to endorse parental legitimacy and their own obligation to obey both contemporaneously and in subsequent years. Specifically, adolescents who perceived their parents to be supportive were more likely to believe that parents had legitimate authority over personal, prudential, and multidomain issues. They were more likely to feel obliged to obey with regard to issues in the opposite-sex relations and multidomain domains. Adolescents who perceived their parents to be high in monitoring were more likely to endorse legitimacy and obligation to obey for issues in the parental expectations and multidomain domains and more likely to feel obliged to obey in the prudential domain as well.

The finding that good parenting predicts adolescents' beliefs about parental legitimacy highlights the interrelationship of parent and adolescent attributes that have been prominent in recent developmental research (Crouter & Head, 2002). Support and monitoring are two components of the constellation of parent attributes associated with the authoritative parenting style. The present findings and those linking adolescents' beliefs with adolescents' disclosure and obedience (Darling et al., 2007) are consistent with the hypothesis that adolescents' legitimacy beliefs are a central link between authoritative parenting and positive adolescent outcomes (Darling & Steinberg, 1993). Further, authoritative parents differ from their authoritarian and permissive peers in making clearer distinctions between personal issues and other domains when regulating adolescent behavior (Smetana, 1995). Parents who are authoritarian (high in behavioral control and intrusive psychological control but relatively low in support) are more likely to set rules in the personal domain, tending to reason about them in either moral or conventional terms. Recent research suggests that adolescents whose parents attempt to regulate issues adolescents define as personal report more internalizing symptoms (Hasebe et al., 2004) and see their parents as more psychologically controlling (Smetana & Daddis, 2002). Thus, adolescents who perceive their parents to be supportive and high in monitoring are more likely to believe parents to be legitimate authorities and to believe that they are obliged to obey them. Adolescents' beliefs predict self-reported adolescent obedience (Darling et al., 2007), voluntary disclosure of information, and perceived parental knowledge (Darling et al., 2006; Smetana et al., 2006). The latter two processes are associated with positive outcomes during adolescence (Crouter & Head, 2002). In addition, because adolescents who perceive their parents to be supportive or high in monitoring are also more likely to believe that parents are legitimate authorities, they are less likely to perceive parents as intruding into their personal jurisdiction.

Adolescent attributes are also important predictors of beliefs. Adolescents who were initially more involved in problem behavior were less likely to believe that parents had legitimate authority over issues in all domains and were less likely to believe that they were obliged to obey parents when they disagreed. These individual differences in problem behavior at the beginning of the study predicted differences in the endorsement of parental legitimacy and obligation to obey both contemporaneously and at each subsequent wave of data collection. Previous research had revealed a relatively small number of adolescents (16% during early and middle adolescence, 9% during late adolescence) who were unlikely to endorse parental authority over issues in the personal, prudential, or multidomain domains (Cumsille et al., 2006). The present findings suggest that relatively high involvement in problem behavior may be one correlate of rejection of parental authority—both as legitimate and as requiring obedience. This co-occurrence of involvement in problem behavior and rejection of authority is consistent with much of what is known of the etiology of chronic adolescent delinquency (e.g., Jessor & Jessor, 1977; Moffitt, 1993; Patterson et al., 1989) and research on adolescent substance users' belief that substance use is a matter of personal discretion (Nucci et al., 1991). It should be noted, however, that prior research on adolescents' attitudes toward authority and delinquency has focused on rebelliousness or rejection of authority as a global or personality characteristic. When examined at the issue level, however, there is evidence that the voluntary disclosure of adolescents who are involved in disapproved activities is more sensitive to their own beliefs about their obligation to obey and to parental rules than is that of their peers (Darling et al., 2007). Thus, although adolescents who are relatively more involved in problem behavior are less likely to endorse parental authority or obligation to obey, they may nonetheless believe that parents are legitimate authorities in some areas and these beliefs have consequences. The direction of the causal relationship between involvement in problem behavior and adolescents' beliefs remains unclear.

This research is also unique in examining longitudinal changes in both legitimacy beliefs and beliefs in adolescents' obligation to obey. Research on normative change in adolescents' and parents' beliefs has primarily focused on beliefs about the legitimacy of parental authority. This is a natural extension of the theoretical underpinnings of social-cognitive domain theory and the processes leading to developmental change in parent–adolescent conflict (e.g., Smetana, 1999). More recently, however, researchers have begun to focus on individual differences in adolescents' beliefs and the consequences of individual differences for adolescent disclosure (Darling et al., 2006; Smetana et al., 2006) and obedience (Darling et al., 2007). Although legitimacy beliefs and beliefs in the obligation to obey are interrelated, they are not identical either conceptually or empirically. For example, U.S. adolescents who were interviewed about their reasoning about disobedience often described areas that parents can (or even should) try to control but which they do not feel obliged to obey (data available upon request). Romantic relations, where they go with friends, substance use, and other issues that combine facets of the prudential and the personal are often cited as examples of issues where adolescents acknowledge their parents' concerns but believe that their own judgment supersedes their obligation to obey. As researchers begin to try to understand the active role that adolescents play in their own development, in shaping parent – adolescent relationships, and in the socialization process, adolescents' beliefs and reasoning about their obligation to obey parents may emerge as a key concept.

Although adolescents whose perceptions of parents and problem behavior differ at the start of the study also differ in beliefs about legitimacy and obligation to obey, the developmental trajectories of their beliefs do not appear to. There is very limited evidence that change in beliefs over time varies with either initial perceptions of parenting or problem behavior. This suggests a model of normative change in adolescents' beliefs such that children enter adolescence differing in their likelihood of endorsing parental authority and obligation to obey and become decreasingly likely to do so as they move toward adulthood. In other words, all adolescents grow less likely to endorse parental authority as they get older, but some adolescents begin that decline believing that parents have the right to set rules about many areas of their lives, and others enter adolescence believing that their parents have a right to set rules about few. For example, the predicted likelihood of endorsing legitimacy by a 12-year-old adolescent who experiences low support and monitoring and is involved in problem behavior is roughly equivalent to the predicted likelihood of a 17-year-old adolescent uninvolved in problem behavior who experiences high support and monitoring. Investigation of the consequences of the premature assertion of personal jurisdiction over areas normatively defined as legitimately within parental control may offer a promising avenue for distinguishing between pseudo- and mature autonomy.

Should this work be replicated and extended, lack of variation in the decline in adolescents' beliefs in the legitimacy of parental authority and obligation to obey would be important in what it tells us about the developmental course of individual differences in adolescents' beliefs. In particular, a limited body of cross-sectional research has documented the association of family climate with individual differences in legitimacy beliefs during adolescence (e.g., Smetana et al., 2006). Cross-sectional research does not, however, allow insight into whether these differences reflect differences that existed in childhood and remained stable throughout adolescence, differences from childhood that grew larger or smaller over time, or differences that emerged only during adolescence. The present research suggests that differences in adolescents' beliefs about legitimacy and obligation to obey have already emerged prior to adolescence and remain stable, despite a normative decline. If so, the focus of future research into the sources of these differences should be on middle childhood and early adolescence rather than on the middle or the late adolescent periods.

The setting for this study, Chile, is somewhat atypical for a study of normative change and individual differences in adolescents' beliefs about parental legitimacy and their own obligation to obey. Because contemporary Chile is in the midst of a societal change perhaps analogous to that of the 1960s in the United States and Western Europe, the extent to which these findings will generalize to other social contexts or historical periods is unclear. For this reason, the consistency of these findings and previous research on normative changes in adolescents' legitimacy beliefs is reassuring. It should be noted that these findings differ from those previously reported in two ways, however. First, prior interview research conducted in Brazil (Nucci, Camino, et al., 1996) revealed social class differences in adolescents' reasoning about parental legitimacy that were not found in this Chilean sample. Although it is possible that this reflects a difference between Chilean and Brazilian society, it is perhaps more likely that these findings reflect methodological differences in how the focal constructs of the studies were assessed. In addition, the antiauthoritarian social climate of contemporary Chile may flatten social differences in adolescents' beliefs, making them more difficult to detect using survey methodology. This study also failed to replicate previously reported differences in adolescents' legitimacy beliefs as a function of family structure (e.g., Smetana, Yau, Restrepo, et al., 1991). It is important to note in this regard that recent changes in the divorce laws make marital status somewhat difficult to assess in Chile. Until 2005, legal divorce could only be obtained through the Catholic Church. It was thus very common for couples to separate and establish long-term relationships and households with new partners while remaining legally married to their first spouse. These cultural practices may muddy the distinction between marital statuses both in terms of measurement and in the way that adolescents think about marital status, thus contributing to our null findings.

The present study is limited in several respects, most importantly in that it relies solely on adolescents' self-reports, limiting causal inference. Although self-report may be the most accurate way to measure adolescents' beliefs and perceptions of parents, objective measures of parent behavior and adolescents' involvement in problem behavior would allow us to make clearer inferences. The scarcity of significant interactions between time and either perceived parenting or problem behavior, however, suggests that bias associated with concurrent reports of predictors and outcomes are minimized in these data. In addition, because HGLM controls for between-adolescent differences when modeling within-adolescent differences, reporter biases are also minimized at this level. These data are also limited by the dichotomous measurement of adolescents' beliefs about the legitimacy of parental authority and obligation to obey. Although this was a necessary compromise to facilitate gathering data in a large and varied sample, it might not fully reflect the complex beliefs that adolescents are likely to hold about parental authority. The consistency of these results with prior findings using in-depth interviews (e.g., Smetana, 1988) and surveys that allowed for multicategorical responses (e.g., Smetana & Chuang, 2001) are thus particularly reassuring. Finally, as is true of most longitudinal studies, self-selection bias is likely to have made this sample better functioning than the population of adolescents as a whole. However, this bias will tend to reduce variability in the measures, thus making our tests more conservative than they would be in a more representative sample.

Adolescence is a liminal period when youth are in the process of establishing autonomous self-regulation at the same time that many parents are still actively trying to regulate their behavior. During this period, parents may continue to monitor their adolescents and set rules for them, but the extent to which those efforts are successful depends upon the adolescents themselves. It is unlikely, for example, for adolescents to be offered drinks when their parents are standing next to them. Although adolescents' beliefs that parents can legitimately regulate their behavior and that they are obliged to obey parents' rules decline as they move toward adulthood, the starting point of that downward trajectory may partly depend upon their perception of parents and the success of socialization during childhood and early adolescence. The extent to which adolescents' beliefs about parental legitimacy and obligation to obey can be reestablished, once lost, is unclear. Individual differences in adolescents' beliefs about parental legitimacy and obligation to obey highlight the difference between rejection of parental oversight and the development of mature autonomy (Darling et al., 2005). Recent work on the central role adolescents play in the ability of parents to monitor effectively (Crouter & Head, 2002) has focused attention on the systemic nature of what have been called "parenting" processes and the active role that adolescents play in them. These results suggest that adolescents' beliefs may be an important part of that system.

References

Barber, B. K., Stolz, H. E., & Olsen, J. A. (2005). Parental support, psychological control, and behavioral control: Assessing relevance across time, culture, and method. *Monographs of the Society for Research in Child Development*, *70*(4).

Collins, W. A. (1992). Parents' cognitions and developmental changes in relationships during adolescence. In I. E. Sigel & A. V. McGillicuddy-DeLisi (Eds.), *Parental belief systems: The psychological consequences for children* (2nd ed., pp. 175–197). Hillsdale, NJ: Lawrence Erlbaum.

Crouter, A. C., Bumpus, M. F., Davis, K. D., & McHale, S. M. (2005). How do parents learn about adolescents' experiences? Implications for parental knowledge and adolescent risky behavior. *Child Development*, *76*, 869–882.

Crouter, A. C., & Head, M. R. (2002). Parental monitoring and knowledge of children. In M. H. Bornstein (Ed.), *Handbook of parenting* (Vol. 3, pp. 461–484). Mahwah, NJ: Lawrence Erlbaum.

Cumsille, P., Darling, N., Flaherty, B. P., & Martínez, M. L. (2006). Chilean adolescents' beliefs about the legitimacy of parental authority: Individual and age-related differences. *International Journal of Behavioral Development*, *30*, 97–106.

Cumsille, P., Darling, N., & Peña-Alampay, L. (2002, April). *Legitimacy beliefs and parent-adolescent conflict and adjustment in adolescence: A Chilean and Filipino comparison*. Poster session presented at the biennial meeting Society for Research on Adolescent Development, New Orleans, LA.

Darling, N., Cumsille, P., Caldwell, L. L., & Dowdy, B. (2006). Predictors of adolescents' disclosure strategies and perceptions of parental knowledge. *Journal of Youth and Adolescence*, *35*, 667–678.

Darling, N., Cumsille, P., & Martínez, M. L. (2007). Adolescents as active agents in the socialization process: Legitimacy of parental authority and obligation to obey as predictors of obedience. *Journal of Adolescence, 30*, 297–311.

Darling, N., Cumsille, P., & Peña-Alampay, L. (2005, Summer). Rules, legitimacy of parental authority, and obligation to obey in Chile, the Philippines, and the United States. *New Directions for Child and Adolescent Development*, *108*, 47–60.

Darling, N., & Steinberg, L. (1993). Parenting style as context: An integrative model. *Psychological Bulletin, 113*, 487–496.

Fuligni, A. J. (1998). Authority, autonomy, and parent-adolescent conflict and cohesion: A study of adolescents from Mexican, Chinese, Filipino, and European backgrounds. *Developmental Psychology, 34*, 782–792.

Hasebe, Y., Nucci, L., & Nucci, M. S. (2004). Parental control of the personal domain and adolescent symptoms of psychopathology: A cross-national study in the United States and Japan. *Child Development*, *75*, 815–828.

Instituto Nacional de la Juventud. (1999). *Familia y vida privada de los jóvenes (No. Análisis y difusión de la segunda encuesta nacional de juventud: Cuadernillo Tematico 4).* Santiago de Chile: Instituto Nacional de la Juventud.

Jessor, R., & Jessor, S. (1977). *Problem behavior and psychosocial development: A longitudinal study of youth*. New York: Academic Press.

Kerr, M., & Stattin, H. (2000). What parents know, how they know it, and several forms of adolescent adjustment: Further support for a reinterpretation of monitoring. *Developmental Psychology, 36*, 366–380.

Kerr, M., Stattin, H., & Trost, K. (1999). To know you is to trust you: Parents' trust is rooted in child disclosure of information. *Journal of Adolescence, 22*, 737–752.

Martínez, M. L., Cumsille, P., & Thibaut, C. (2006). Chile. In J. J. Arnett (Ed.), *International encyclopedia of adolescence* (Vol. 2, pp. 167–178). Oxford, England: Routledge.

Moffitt, T. E. (1993). Adolescent-limited and life-course-persistent antisocial behavior: A developmental taxonomy. *Psychological Review, 100*, 674–701.

Nucci, L. P., Camino, C., & Sapiro, C. M. (1996). Social class effects on northeastern Brazilian children's conceptions of areas of personal choice and social regulation. *Child Development*, *67*, 1223–1242.

Nucci, L. P., Guerra, N., & Lee, J. (1991). Adolescent judgments of the personal, prudential, and normative aspects of drug usage. *Developmental Psychology, 27*, 841–848.

Nucci, L. P., Killen, M., & Smetana, J. G. (1996). Autonomy and the personal: Negotiation and social reciprocity in adult-child social exchanges. *New Directions for Child Development*, *73*, 7–24.

Parke, R. D., & Buriel, R. (1998). Socialization in the family: Ethnic and ecological perspectives. In N. Eisenberg (Ed.), *Handbook of child psychology* (Vol. 3, pp. 463–546). New York: Wiley.

Patterson, G. R., DeBaryshe, B. D., & Ramsey, E. (1989). A developmental perspective on antisocial behavior. *American Psychologist*, *44*, 329–335.

Pettit, G. S., Laird, R. D., Dodge, K. A., Bates, J. E., & Criss, M. M. (2001). Antecedents and behavior-problem outcomes of parental monitoring and psychological control in early adolescence. *Child Development*, *72*, 583–598.

Programa de Naciones Unidas para el Desarollo. (2002). *Desarrollo humano en Chile: Nosotros los Chilenos: Un desafio cultural.* Retrieved May 17, 2006, from http://www.desarollohumano.cl/eleccion2002.htm

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Newbury Park, CA: Sage.

Smetana, J. G. (1988). Adolescents' and parents' conceptions of parental authority. *Child Development, 59*, 321–335.

Smetana, J. G. (1995). Parenting styles and conceptions of parental authority during adolescence. *Child Development*, *66*, 299–316.

Smetana, J. G. (1996). Adolescent-parent conflict: Implications for adaptive and maladaptive development. In D. Cicchetti & S. L. Toth (Eds.), *Adolescence: Opportunities and challenges. Rochester symposium on developmental psychopathology* (Vol. 7, pp. 1–46). Rochester, NY: University of Rochester Press.

Smetana, J. G. (1997). Parenting and the development of social knowledge reconceptualized: A social domain analysis. In J. E. Grusec & L. Kuczynski (Eds.), *Parenting and children's internalization of values: A handbook of contemporary theory* (pp. 162–192). New York: Wiley.

Smetana, J. G. (1999). Context, conflict, and constraint in adolescent-parent authority relationships. In M. Killen & D. Hart (Eds.), *Morality in everyday life: Developmental perspectives* (pp. 225–255). New York: Cambridge University Press.

Smetana, J. G. (2000). Middle-class African American adolescents' and parents' conceptions of parental authority and parenting practices: A longitudinal investigation. *Child Development*, *71*, 1672–1686.

Smetana, J. G. (2002). Culture, autonomy, and personal jurisdiction in adolescent-parent relationships. In R. V. R. Kail & Hayne W. (Eds.), *Advances in child development and behavior* (Vol. 29, pp. 51–87). San Diego, CA: Academic Press.

Smetana, J. G., & Chuang, S. (2001). Middle-class African American parents' conceptions of parenting in early adolescence. *Journal of Research on Adolescence, 11*, 177–198.

Smetana, J. G., Crean, H. F., & Campione-Barr, N. (2005, Summer). Adolescents' and parents changing conceptions of parental authority. *New Directions for Child and Adolescent Development*, *108*, 31–46.

Smetana, J. G., & Daddis, C. (2002). Domain-specific antecedents of parental psychological control and monitoring: The role of parenting beliefs and practices. *Child Development*, *73*, 563–580.

Smetana, J. G., Daddis, C., & Chuang, S. S. (2003). "Clean your room!" A longitudinal investigation of adolescent-parent conflict and conflict resolution in middle-class

African American families. *Journal of Adolescent Research, 18*, 631 – 650.

Smetana, J. G., Metzger, A., Gettman, D. C., & Campione-Barr, N. (2006). Disclosure and secrecy in adolescent-parent relationships. *Child Development, 77*, 201 – 217.

Smetana, J. G., Yau, J., & Hanson, S. (1991). Conflict resolution in families with adolescents. *Journal of Research on Adolescence, 1*, 189 – 206.

Smetana, J. G., Yau, J., Restrepo, A., & Braeges, J. L. (1991). Adolescent-parent conflict in married and divorced families. *Developmental Psychology, 27*, 1000 – 1010.

Snijders, T., & Bosker, R. (1999). *Multilevel analysis*. Thousand Oaks, CA: Sage.

Steinberg, L., & Silk, J. S. (2002). Parenting adolescents. In M. H. Bornstein (Ed.), *The handbook of parenting* (Vol. 1, pp. 133 – 134). Mahwah, NJ: Lawrence Erlbaum.

Steinberg, L., & Silverberg, S. (1986). The vicissitudes of autonomy in early adolescence. *Child Development, 57*, 841 – 851.

Thomas, A., & Chess, S. (1968). *Temperament and behavior disorders in children.* New York: New York University Press.

Turiel, E. (1983). *The development of social knowledge: Morality and convention.* Cambridge, England: Cambridge University Press.

Yau, J., & Smetana, J. (2003). Adolescent-parent conflict in Hong Kong and Shenzhen: A comparison of youth in two cultural contexts. *International Journal of Behavioral Development, 27*, 201 – 211.

Child Development, July/August 2008, Volume 79, Number 4, Pages 1119–1136

The Breadth of Peer Relationships Among Preschoolers: An Application of the Q-Connectivity Method to Externalizing Behavior

Laura D. Hanish, Carol Lynn Martin, Richard A. Fabes, and Hélène Barcelo
Arizona State University

Preschoolers' ($M_{\text{age}} = 48.7$ months) social breadth (the number of peer children interact with and the frequency of these interactions) and the relations between breadth and externalizing behaviors were examined using the Q-connectivity method. After considering age, sex, and classroom effects on social breadth, children's externalizing behaviors were studied in relation to levels of social breadth across variations in frequency of exposure to peers and to patterns of decline in social breadth. Externalizing preschoolers occasionally interacted with most of their classmates. However, their individual peer networks moderately to rapidly became selective, and they engaged in repeated interactions with a relatively small number of peers. Although there were no sex effects, significant age and classroom effects were obtained.

One of the most interesting, but least studied, aspects of children's peer relationships is the substantial variation in the breadth of children's peer affiliations. Some children interact broadly with many peers, whereas others are more selective. To complicate matters, relationship patterns change over time: Friendships form, persist for short or long periods, and often dissolve, and peers move into or out of social networks. Consequently, there is both stability and fluidity, as well as marked individual differences, in children's peer affiliations (Cairns, Leung, Buchanan, & Cairns, 1995; Estell, Cairns, Farmer, & Cairns, 2002; Martin, Fabes, Hanish, & Hollenstein, 2005). It is this variability in children's social breadth (i.e., the number of peers that each child interacts with and the frequency with which each child interacts with every peer) that we highlight in the present study. To do so, we introduce the *Q-connectivity method* (Barcelo & Laubenbacher, 2005) to calculate preschoolers' social breadth. The Q-connectivity method is a variation on social network methods that involves computing the frequency with which each child interacts with every other peer and using this information to build each child's social network at increasing levels of exposure to peers.

We expected to find individual differences in children's social breadth and we hypothesized that variations in social breadth would be related to children's externalizing behavior. We focused specifically on children's externalizing characteristics because, as research on the heterogeneity of aggressive children has highlighted (Rodkin, Farmer, Pearl, & Van Acker, 2000), the fact that some aggressive children maintain successful peer relationships, whereas others have more difficulty speaks to the need to better understand the peer contexts of these children. This issue is directly relevant to the need to better understand peer contagion processes. The conditions that are necessary for peer contagion to occur remain to be documented; for example, it is not well understood whether externalizing behaviors can be spread with just a few contacts with an aggressive child or whether extended interactions are required. Examining the social breadth of externalizing children is important to this area of study because it represents their potential range of influence.

Developmental–Contextual Framework for Conceptualizing the Relations Between Social Breadth and Externalizing Behaviors in Early Childhood

Social breadth refers to the range of peer affiliates—that is, the number of peers that each child interacts with—and the frequency with which each child interacts with every peer. As such, social

---

This research was supported, in part, by grants from the National Institute of Child Health and Human Development awarded to Laura D. Hanish, Carol Lynn Martin, Richard A. Fabes (1 R01 HD45816-01A1); from the National Security Agency (H98230-05-1-0256) awarded to Hélène Barcelo; and from the National Science Foundation awarded to all the authors (0338864). Research support for this project was also provided by the Cowden Endowment Fund. The authors would like to thank the students who contributed to this project and the children and teachers for their participation. Additionally, special thanks go to Mike Dodd for software development and to Mark Reiser for statistical consultation.

Correspondence concerning this article should be addressed to Laura D. Hanish, Program in Family and Human Development, School of Social and Family Dynamics, Arizona State University, Box 873701, Tempe, AZ 85287-3701. Electronic mail may be sent to laura.hanish@asu.edu.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7904-0020

breadth represents a construct that is distinct from friendship, social status, or membership in social networks because it reflects the entire social landscape of children's peer affiliations and is not limited to a particular type or quality of social interaction, although it may relate to these factors. The roots of this construct can be seen in the early work of Challman (1932) and Waldrop and Halverson (1975).

We speculate that social breadth is an organizing feature of peer relationships across the developmental life course. We also speculate that the meaning of social breadth will vary according to developmental level. For example, early in life, social breadth may reflect increasing maturity in social development. Evidence for this supposition comes from research that has shown that, across early childhood, children evidence increasingly greater preference for and competence in engaging their peers (Rubin, Bukowski, & Parker, 1998). Moreover, preschoolers' social networks become larger, denser, and more organized with age (Strayer & Santos, 1996; Vespo, Kerns, & O'Connor, 1996). Additionally, during this time, parents and teachers expect increasing sociability, and children respond to these expectations and pressures by seeking out peer interactions. However, there is still considerable variability and change in these social tendencies during early childhood, making it an ideal time to study the correlates associated with the evolution of social breadth.

The early childhood period is a particularly important developmental period to consider sex differences in social breadth. Research has suggested that, beginning around kindergarten age, boys maintain larger groups than girls. However, among preschoolers, there are fewer clear sex differences in group size, and both boys and girls tend to interact with peers in small groups (Ruble, Martin, & Berenbaum, 2006). These findings, however, pertain to the size of friendship networks and may or may not translate to social breadth. Thus, it is necessary to describe the extent to which there are sex differences in social breadth in early childhood.

It is important to note that social breadth is a construct that does not reside entirely within the child. Rather, children's social breadth reflects their own characteristics as well as those of their peers. Thus, the degree to which children exhibit greater or lesser social breadth may be as much dependent upon the mix of available peers and the match between their own characteristics and behavioral styles and those of their peers as it is an indicator of their underlying propensity to engage with others. This means that children's tendencies to engage with peers may vary from classroom to classroom, depending upon the classroom peer culture; consequently, we would expect to find classroom-related variations in social breadth as well as individual differences in social breadth that are related in meaningful ways to children's characteristics.

Looking beyond contextual and demographic characteristics, we presume that children's externalizing behaviors also contribute to social breadth. Externalizing behaviors are particularly salient in young children's social lives, and it is not surprising that the peer relationships of aggressive children tend to be sparser and less stable over time. For instance, van den Oord, Rispens, Goudena, and Vermande (2000), who collected sociometric nomination data with over 1,000 preschool children, found that children rated by teachers as having more problem behaviors (e.g., aggression, attention problems, and fear) were less likely to be members of cliques and had peer relationships that were less transitive (i.e., children were disliked by their friends' friends). Similarly, Snyder, Horsch, and Childs (1997) found that aggressive preschoolers, compared to nonaggressive preschoolers, developed fewer relationships with peers that were stable over time. Additionally, studies have demonstrated that aggressive children have lower social preference scores and are more likely to be rejected by peers than their nonaggressive counterparts (Snyder et al., 1997; Olson, 1992). Thus, we expected social breadth to be negatively related to externalizing behaviors, including physical aggression, attention problems, and anger, among young children.

The Use of the Q-Connectivity Method to Measure Social Breadth

To measure social breadth, we used a recently developed variation on social network methods called the *Q*-connectivity method (Barcelo & Laubenbacher, 2005). This is a mathematical method of modeling complex systems and their dynamics, and it is derived from the *Q*-connectivity theory for simplicial complexes and graphs (in which the graphs are diagrams with vertices and edges). The method has previously been applied in nonsocial areas, such as search-and-rescue operations (Barcelo & Laubenbacher, 2005). From a social developmental perspective, the *Q*-connectivity method makes it possible to measure aspects of peer relationships that have been difficult to study using existing methods because it allows for an assessment of the changing structure of individual children's interaction patterns with all available peers. This is made possible by *Q*, which represents exposure; thus, as *Q* increases, exposure to peers may also increase.

The *Q*-connectivity method is marked by two unique features that make this a significant innovation for the study of peer relationships. First, the *Q*-connectivity method maps networks individually for each child in the sample rather than for a group of children—for example, classmates—as a whole as is often done with traditional social network procedures (Cairns, Cairns, Neckerman, Gest, & Gariepy, 1988; Richards & Rice, 1981). This is a useful distinction to make because when networks are mapped for a group as a whole, information about networks of those children who are isolated or peripheral is sacrificed in exchange for a parsimonious representation of the social structure of the group as a whole. Thus, by using the *Q*-connectivity method, each child's unique patterns of interaction with *all* available peers are assessed.

Second, the *Q*-connectivity method capitalizes on the dynamic nature of peer interactions by repeatedly assessing the structure of children's peer interactions at differing levels of *Q*, thus providing an estimate of children' exposure to every peer. In the present study, we relied on repeated observations of children's peer interaction partners obtained over the course of a semester. This is an important feature of the *Q*-connectivity method because it provides a means for measuring the fluctuations that occur in children's peer interaction partners across a period of time. In comparison, traditional social network methods most often rely on data obtained using self-reports or sociometric nominations collected at one discrete time; thus, the resulting social networks represent aggregated views of relationships that provide relatively static pictures of children's affiliation patterns (for recent exceptions to this, see Moody, McFarland, & Bender-deMoll, 2005; Snijders, 2005).

By applying the *Q*-connectivity method to the study of social breadth, we are able to calculate multiple measures of social breadth for each child that change as a function of exposure to peers (see the Method section for details about how this is operationalized). On average, children should have a few interactions (low exposure) with many peers but more frequent interactions (high exposure) with fewer peers. By measuring social breadth at lower and higher levels of exposure, we are able to (a) examine variations in the structure and meaning of social breadth and (b) assess the drop-off of social breadth as we move across a continuum of how often children were seen playing together.

The Present Study

We described preschoolers' social breadth along a dimension representing the relative exposure to peers and examined this in relation to three aspects of externalizing behavior—namely, physical aggression, attention problems, and anger. We tested two sets of hypotheses regarding the relations between externalizing behaviors and social breadth. Both sets of hypotheses take into consideration the drop-off of social breadth as the criterion for peer exposure is increased. However, this drop-off was conceptualized differently for both hypotheses. This allowed us to explore the relations between externalizing behaviors and (a) variations in *social breadth scores* along a dimension of increased peer exposure (Hypothesis 1) and (b) variations in the *patterns of decline* (trajectories) of social breadth scores as peer exposure increased (Hypothesis 2).

Hypothesis 1. We examined the relations between externalizing behavior and social breadth at each of several consecutive levels of peer exposure. In general, we expected to find a negative relation between externalizing behaviors (physical aggression, attention problems, and anger) and social breadth, which would indicate that externalizing children interact with a smaller set of peers than do nonexternalizing children. This hypothesis is supported by previous studies that have shown negative relations between externalizing behavior and other indicators of the quality of young children's peer relationships (Snyder et al., 1997; van den Oord et al., 2000). However, we also expected that the relations would be weak and nonsignificant at low exposure because these reflect relatively short-lived peer relationships (i.e., children are only required to interact a small number of times to be counted) but significantly negative at medium to high levels of exposure because these reflect more enduring peer relationships (i.e., children are required to interact more frequently to be counted). Such a finding would support the idea that externalizing children have relationships with peers that are less stable than those of nonexternalizing children (Snyder et al., 1997).

Hypothesis 2. We also considered the possibility that there are systematic variations in the patterns of drop-off of social breadth—that children can be distinguished according to their trajectories of social breadth as peer exposure increases. For instance, some children may have a relatively high intercept, followed by a slowly declining slope, whereas others may have a relatively high intercept followed by a steeply declining slope. Still others may have a relatively low intercept with trajectories that similarly vary in steepness. We expected that these variations in patterns could be meaningfully explained by children's externalizing behaviors. An implicit assumption underlying this idea is that a person-oriented approach can be useful in identifying and categorizing variations in patterns among children (Magnusson, 1998). This type of approach has been successfully used in prior research on children's peer relationships to describe and explain the heterogeneity in children's social relationships (Rodkin et al., 2000).

We hypothesized that children could be clustered according to systematic variations in their trajectories of social breadth and that these clusters would be meaningfully differentiated by both intercepts and slopes. In particular, we expected to find one group evidencing a moderate to high intercept (which would indicate moderate to high social breadth at low peer exposure) with a steeply declining slope as exposure increases (which would indicate low stability of peer relationships). We further expected that this pattern of social breadth would characterize the children who exhibit relatively high rates of externalizing behaviors. This hypothesis is consistent with prior research that has demonstrated that the quality and duration of externalizing children's peer relationships tend to be lower than those of more competent children (Snyder et al., 1997).

Additional research questions. We also considered the relations between age and sex and social breadth, and we controlled for their effects when relevant. We expected to find significant age differences in social breadth (Strayer & Santos, 1996; Vespo et al., 1996). But previous research on sex differences in young children's peer group size has been equivocal (Ruble et al., 2006); therefore, this was tested as an exploratory hypothesis. Although prior research has been ambiguous about the likelihood of sex differences in social breadth, there is empirical support for the hypothesis that sex moderates the relation between externalizing behaviors and peer socialization processes in young children (Hanish, Martin, Fabes, Leonard, & Herzog, 2005; Zimmer-Gembeck, Geiger, & Crick, 2005). Thus, we tested for moderation by sex. Doing so is important to current efforts to understand how differences in girls' and boys' physical aggression translate to the peer culture. Finally, we also considered classroom influences. With six classrooms in this sample, we did not have a classroom sample size that was sufficiently large to fully explore the contribution of classrooms using a multilevel design. However, we were able to address this issue by testing for the amount of variance explained by classroom level variables and by controlling for classroom as a random effect.

Method

Participants

Participants were students and teachers from six classrooms in three preschools in an urban area in the Southwest. All the schools provided full-day child care. Class sizes ranged from 17 to 20. All members of each class were recruited for participation in the study; the parental permission rate was 98%. Participants were recruited across two waves over two consecutive years. Twenty-seven children participated in both waves of data collection; these children participated as target children only during their 1st year, but they were counted as peers whenever they were observed to interact with a target child during the 2nd year. The sample consisted of 97 children with complete data (complete data were not available for 8 children), with age ranging from 37 to 64 months ($M = 48.7$, $SD = 6.1$). There were similar numbers of boys (52%) and girls. Children were predominantly non-Hispanic White (55%) and Hispanic White (32%, predominantly Mexican American), with the remainder Asian American (6%), Native American (3%), and African American (3%). Family income ranged from under $10,000 to over $140,000 (median $50,000–$60,000). Almost two thirds of the children lived in two-parent families (64%).

Procedures and Measures

For the present study, parents reported on child and family demographic variables and teachers reported on children's externalizing behaviors in the classroom. Behavioral observations of children's peer interactions in the classroom were used as the basis for calculating social breadth scores.

Observations of peer interactions. During the fall semester, trained independent observers used a scan observation procedure to record children's peer interaction partners (Hanish, Martin, et al., 2005; Martin & Fabes, 2001). Observers were in children's classrooms daily ($n = 56$ days in Wave 1 and $n = 69$ days in Wave 2) during the hours in which most children were awake and in attendance (e.g., between 9 a.m. and 4 p.m., excluding nap time), and they observed children during structured (e.g., teacher-led story time) as well as during unstructured (e.g., free play) activities in the classroom and on the playground. Observers rotated through a randomly ordered list of children, observing each child for 10 s. Once observers reached the bottom of the list, they waited for approximately 5 min and then began again. Observers completed the entire list during each observation session. Midway through the semester, the list was reversed to control for any possible bias due to order of being observed.

Observations were recorded using handheld computers. A total of 28,726 usable observations were collected. An average of 252 observations per child were obtained (range = 53–484). There were no univariate outliers for the number of observations. Observers recorded whether each child was present and available for coding, present but unavailable (e.g., in the bathroom and in the director's office), or absent. Several codes were recorded for each child, but for the purposes of the present study, only the peer partner codes were relevant. Peers were coded whenever the target child's predominant activity during the 10-s scan involved direct interaction with one or more classmates (e.g., social conversation, rule-based play, interactive play, rough-and-tumble play, and aggression) or parallel play with one or more classmates (i.e., playing alongside one another in the same activity). In preschoolers, parallel play often precipitates, or occurs as a lull in the midst of, bouts of direct peer interaction. Observers recorded as many as five peer partners during each observation. A total of 24.7% of the observations included a single partner, 9.1% included two partners, 3.4% included three partners, 1.2% included four partners, and only 0.8% included five or more partners (observations of more than five peers were very rare; in these instances, the five peers who were most closely involved with the target child were coded); the rest (60.8%) had no peer play partners (i.e., solitary activities or interactions with the teacher).

Prior to beginning data collection, observers were extensively trained on the coding procedures. Reliability was assessed on a regular basis throughout the semester by having two independent observers simultaneously code the same child (obtained on 12.9% of all observations). Reliability estimates for the peer partner codes were high; percent agreement was 91% for number of peer partners and 96% for identity of peer partners. To calculate kappas, we first used a weighting technique to balance nonsquare contingency tables (due to the fact that observers could use different numbers of categories—i.e., peer IDs—in making their ratings; Stein, Devore, & Wojcik, 2005). Kappas were then calculated for data collected in Waves 1 and 2 separately, κs = .91 (95% confidence interval [CI] = 0.88 – 0.93) and .91 (95% CI = 0.89 – 0.93), respectively. These observational procedures have been used in previous studies and have demonstrated very good reliability and validity (Fabes, Shepard, Guthrie, & Martin, 1997; Hanish, Martin, et al., 2005; Martin & Fabes, 2001).

Calculation of social breadth. Social breadth was operationally defined to include all observed peer contacts, and it can be visualized graphically as well as calculated mathematically. We first describe the graphical method for representing social breadth, followed by the way in which we computed the measure. Compared to the graphs that are produced using traditional social network methods (e.g., Cairns et al., 1988; Richards & Rice, 1981), the graphs that are produced using the *Q*-connectivity method are unique along two dimensions. First, rather than producing a single graph for a class as a whole (or other specified group), separate sets of graphs are produced for each individual in the sample. Thus, we obtained 97 sets of graphs, with each set representing an individual child's set of social networks across *Q* (level of exposure). Second, rather than producing just one single graph for each individual, *Q* graphs are produced for each child, where *Q* can vary from 1 to 69 (where 69 = the maximum number of days during the semester in which observations were conducted, i.e., maximum possible exposure).

To facilitate the description of the graphs, we provide an example. In Figure 1, the set of graphs for a target child (identified as Boy 1) is illustrated. Note that the target child is not visually represented in the graphs—only his peers can be seen—instead, the target child is assumed to be above the graph and edges from the target child to the peers are assumed to be present. This decision to exclude the target child from the visual representation of the graph reduces the clutter in the graphs and improves their readability. The vertices represented in the first graph of Figure 1 correspond to the peers who were observed to interact with Boy 1 *on at least 1 day* during the semester (square vertices indicate male peers and circular vertices indicate female peers; the placement and size of the vertices on the graphs are arbitrary). The children identified as 711, 739, and 1023 were among these peers (also see Figure 2, in which the first graph for Boy 1 is enlarged for readability). Counting the number of vertices in this graph reveals that Boy 1 interacted with 13 peers at least once during the semester. There is an edge between two vertices (i.e., between two peers) if the two peers and Boy 1 interacted as a group on at least 1 day. Again, looking at the first graph in Figure 1 (and Figure 2), we see that this was the case, for example, for Peers 739 and 1023 but not so for Peers 711 and 1051. The subsequent graphs are constructed in a similar manner. For example, the vertices in the third graph represent the peers who Boy 1 was observed to interact with *on at least 3 days* during the semester (the third graph for Boy 1 is enlarged in Figure 3). For Boy 1, there were only four such peers: 711, 729, 739, and 1023. In the same manner, there is an edge between two vertices in the third graph if Boy 1 interacted with two peers as a group on at least 3 days during the semester. In Figure 1, graph $Q = 3$, we see that this was the case for 739 and 1023 but not so for 711 and 729 nor for 711 and 1023 (also shown in Figure 3). This means that although Boy 1 interacted with all these peers on at

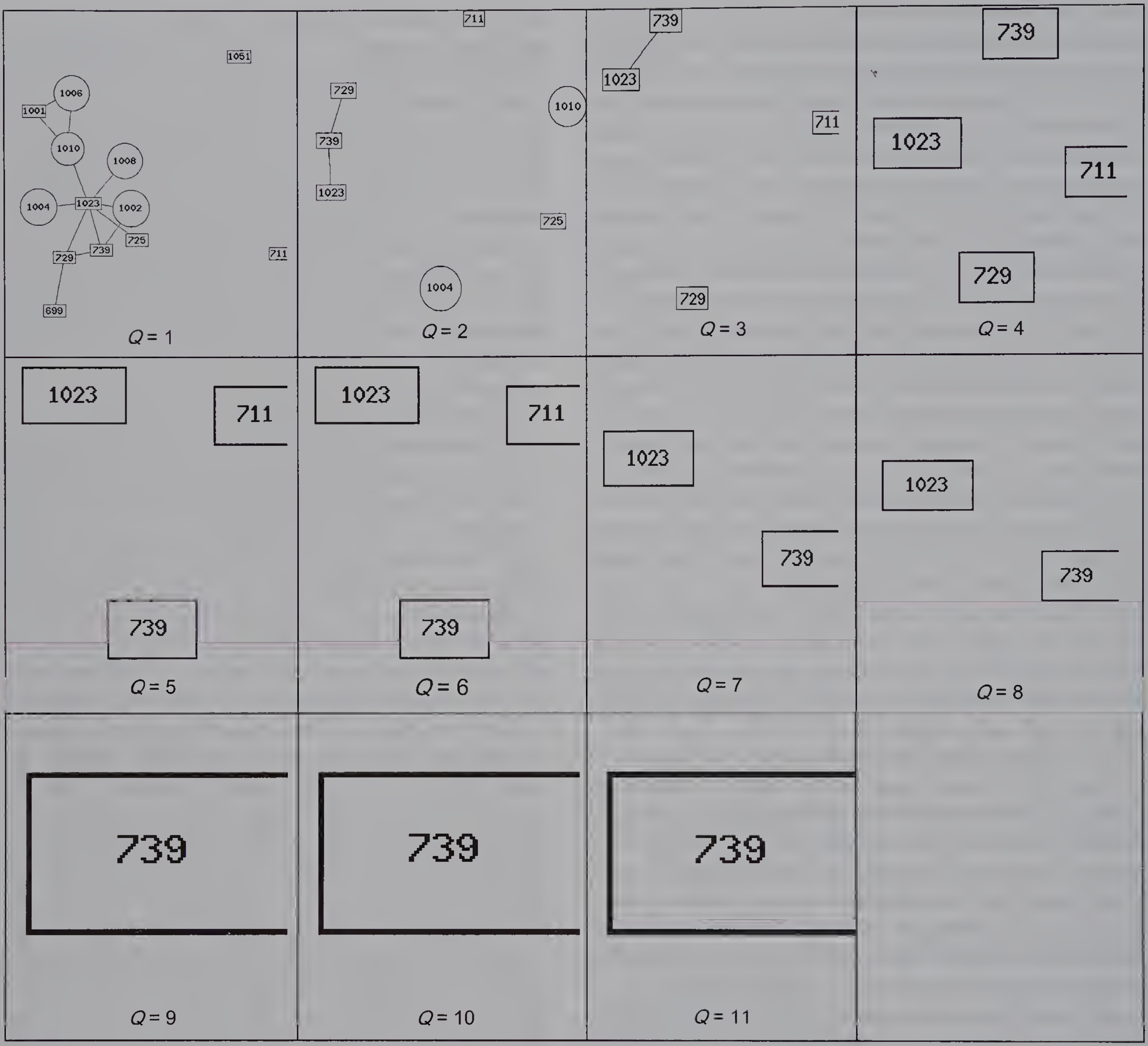

1051
1006
1001
1010
1008
1004
1023
1002
725
729
739
711
699
Q = 1
711
729
1010
739
1023
725
1004
Q = 2
739
1023
711
729
Q = 3
739
1023
711
729
Q = 4
1023
711
739
Q = 5
1023
711
739
Q = 6
1023
739
Q = 7
1023
739
Q = 8
739
Q = 9
739
Q = 10
739
Q = 11

Figure 1. Graphs illustrating social breadth for Boy 1 at $Q = 1$ through $Q = 11$.
Note. This target child was observed with no peers at $Q = 12$ or beyond. Q is operationalized as days (e.g., the value of the function at $Q = 8$ indicates the number of peers observed interacting with a target child on at least 8 days during the semester). Male peers are identified with square vertices and female peers are identified with circular vertices. The edges connecting vertices represent connections among peers and are not considered in the present study.

least 3 days, only Peers 739 and 1023 played at the same time with Boy 1 on at least 3 days. Note that, for Boy 1, after $Q = 11$, no more graphs are shown (see Figure 1). This is because Boy 1 did not interact with any children for more than 11 days during the semester. However, because there were as many as 69 days of observation, all graphs at $Q = 12$ and beyond are empty for Boy 1 and thus not depicted.

The graphs were generated using a data processing program developed for the purpose of facilitating Q-connectivity analyses. The program also outputs several measures that can be calculated from the graphs at each value of Q. The observations of each target child and his or her play partners served as input data. It is important to note, however, that observational data are not required for the use of this method. Any method of data collection that provides repeated assessments of an individual's interaction or relationship partners could be used. Examples of such methods include diary methods, beeper methods, daily self-reports, and phone

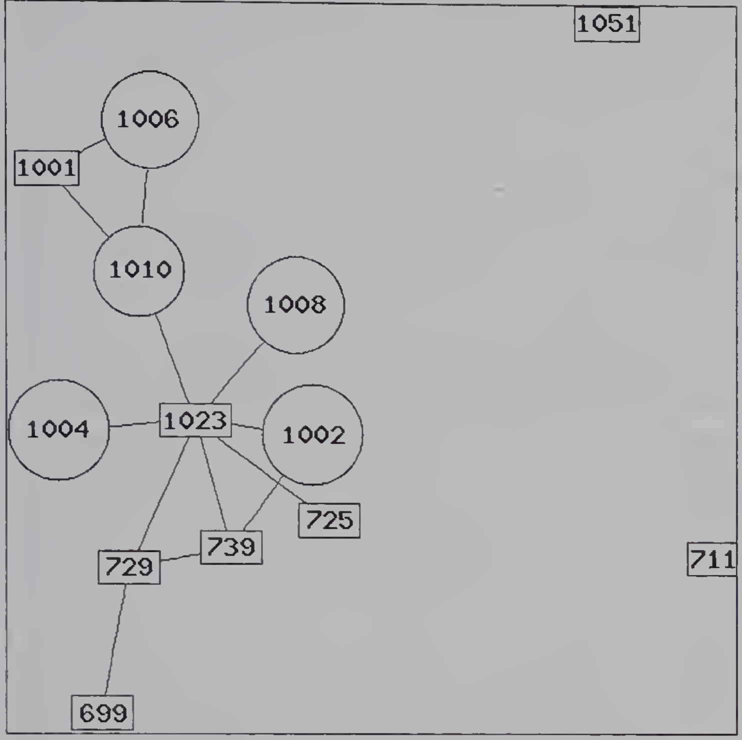


Figure 2. Enlarged $Q = 1$ graph for Boy 1.
Note. Q is operationalized as days (e.g., the value of the function at $Q = 1$ indicates the number of peers observed interacting with a target child on at least 1 day during the semester). Male peers are identified with square vertices and female peers are identified with circular vertices. The edges connecting vertices represent connections among peers and are not considered in the present study. The full set of graphs for this target child is illustrated in Figure 1.

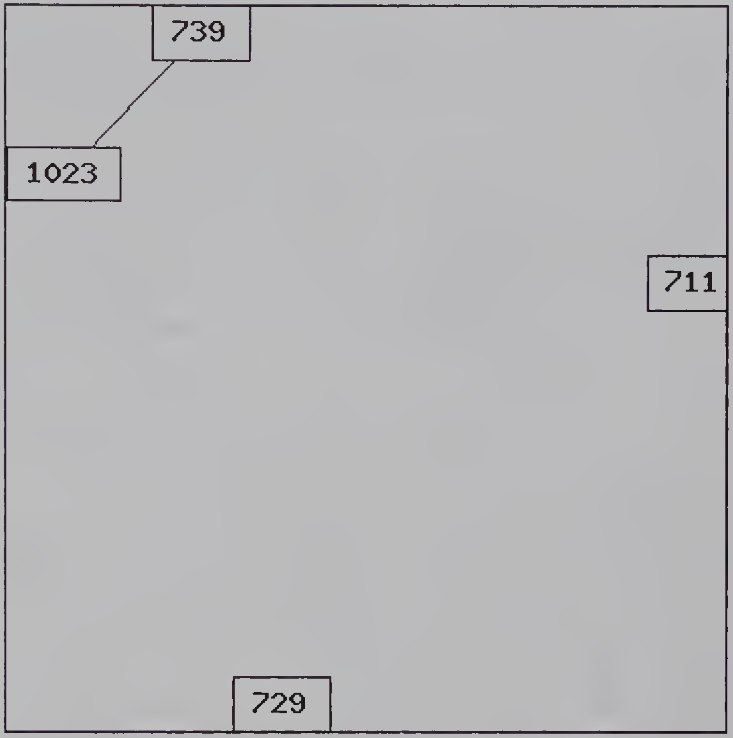


Figure 3. Enlarged $Q = 3$ graph for Boy 1.
Note. Q is operationalized as days (e.g., the value of the function at $Q = 3$ indicates the number of peers observed interacting with a target child on at least 3 days during the semester). Male peers are identified with square vertices and female peers are identified with circular vertices. The edges connecting vertices represent connections among peers and are not considered in the present study. The full set of graphs for this target child is illustrated in Figure 1.

call methods (Asakawa & Csikszentmihalyi, 2000; Birnbaum, Reis, Mikulincer, Gillath, & Orpaz, 2006; McHale, Crouter, & Bartko, 1992). Moreover, recent technological advances, including Bluetooth technology and handheld computers may facilitate multiple assessments of children's peer networks (Eagle & Pentland, 2006; Le, Hat, & Beal, 2006).

There are two important features of the data-processing program. The first is that the program systematically cleans the data by screening out data entry errors and invalid observations. The second is that the program calculates the values needed to quantify and graphically depict children's peer networks. The essential and defining feature of the Q-connectivity method entails conducting simple counts of the number of times that each target child is observed to interact with every possible peer. In the present study, we used the following method to count peers. Because we were specifically interested in using these counts to define social breadth, multiple interactions with the same peer on the same day were counted only once and only one observation with a peer per day was required for the peer to be counted on that given day. It is important to note, however, that the Q-connectivity method is flexible, such that more stringent criteria could be used to identify peers (i.e., increasing the proportion of observations per day that are required before a particular peer is "counted"). Doing so enables one to use this method to assess increasingly closer and more enduring peer relationships, in addition to social breadth. It is also possible to calculate many additional Q-connectivity parameters, which build on the counts of peers just described. For instance, one may count the edges that connect vertices in the graphs. These signify group interactions, and they are described in Hanish et al. (2007). These, and other Q-connectivity parameters, are not described in detail here because of space limitations.

To illustrate our calculations of social breadth, Boy 1, whose graphs are shown in Figure 1, was observed with exactly 13 peers at least once during the semester. His social breadth score was 13 at $Q = 1$. That child was observed with exactly four of those peers at least three times, so he had a score of 4 at $Q = 3$. Finally, because he was observed with exactly one of those peers at least 10 times, he had a score of 1 at $Q = 10$. This pattern of social breadth is shown in Figure 1 by counting the number of vertices (i.e., peers) at each value of Q. In comparison, Figure 4, which represents the peer interactions for Boy 2, depicts a more

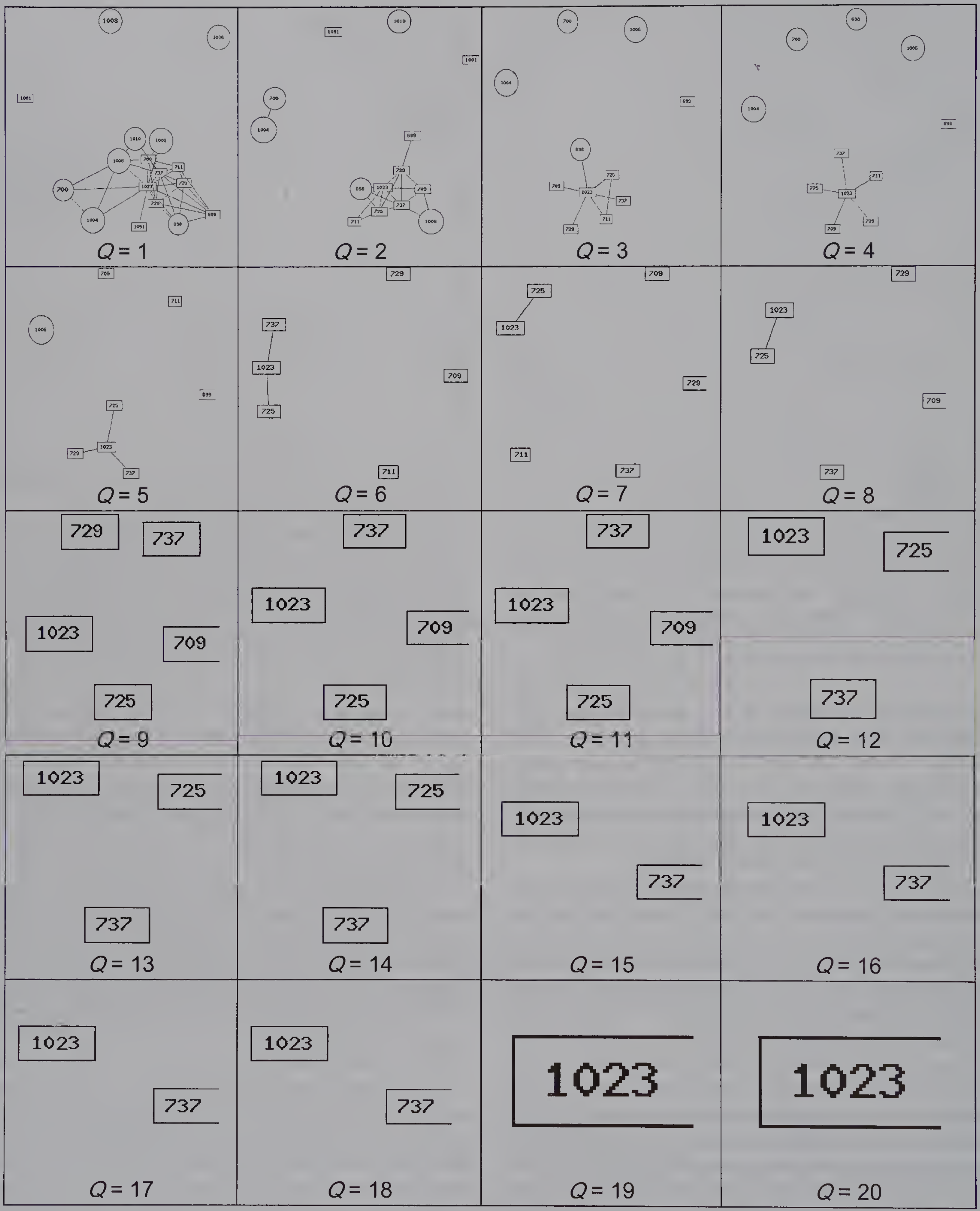

Q = 1
Q = 2
Q = 3
Q = 4
709
711
1006
725
699
1023
729
737
Q = 5
729
737
1023
709
725
711
Q = 6
725
709
1023
729
711
737
Q = 7
729
1023
725
709
737
Q = 8
729
737
1023
709
725
Q = 9
737
1023
709
725
Q = 10
737
1023
709
725
Q = 11
1023
725
737
Q = 12
1023
725
737
Q = 13
1023
725
737
Q = 14
1023
737
Q = 15
1023
737
Q = 16
1023
737
Q = 17
1023
737
Q = 18
1023
Q = 19
1023
Q = 20

Figure 4. Graphs illustrating social breadth for Boy 2 at $Q = 1$ through $Q = 20$.

Note. This target child was observed with Peer 1023 through $Q = 36$. Q is operationalized as days (e.g., the value of the function at $Q = 8$ indicates the number of peers observed interacting with a target child on at least 8 days during the semester). Male peers are identified with square vertices and female peers are identified with circular vertices. The edges connecting vertices represent connections among peers and are not considered in the present study.

extensive pattern of social breadth. At $Q = 1$, this child had a score of 17. His score at $Q = 3$ was 11 and at $Q = 10$ was 4.

Two features of social breadth scores deserve comment. First, according to the method used to define and construct the graphs, as Q (exposure) increases, social breadth scores naturally decrease—that is, as the criterion for the number of times children were seen together increases, the natural function is for there to be fewer peer interactions that meet the criterion. Thus, selectivity in peer exposure is reflected by these changing graph indices. Second, children who were observed playing together for 2 days over the course of the semester could have been observed playing together on two consecutive days or the 2 days could have been separated by several days, weeks, or even months. Thus, as Q increases, social breadth reflects more frequent interactions (greater exposure) with peers, but the length of time between interactions with the same peers is not consistent. For the purpose of this study, we are not concerned with the length of time between observations per se, and we are focused principally on the degree of exposure to peers across the semester.

Teachers' ratings of externalizing behaviors. Teachers completed several measures of classroom externalizing behavior: physical aggression, attention problems, and anger. Physical aggression and attention problems were measured using the Aggressive Behavior (25 items, sample item "This child attacks people") and Attention Problems (9 items, sample item "This child can't concentrate") subscales of the Teacher Report Form of the Child Behavior Checklist (Achenbach & Rescorla, 2000). Each item was completed using a 3-point Likert-type scale. Both scales have been shown to be reliable and valid for preschool children (Achenbach & Rescorla, 2000).

Anger was assessed using a brief subscale of the Child Behavior Questionnaire (CBQ; Goldsmith & Rothbart, 1991; Rothbart, Ahadi, & Hershey, 1994). This measure has been used extensively with young children in prior research to assess negative emotions and regulatory abilities. The items on the CBQ were completed using a 7-point Likert-type scale, with response options ranging from *extremely untrue of this child* to *extremely true of this child*. The Anger subscale consisted of 10 items (sample item was "Gets mad when provoked by other children"; $\alpha = .92$).

Results

After conducting preliminary analyses of social breadth and the externalizing measures, we tested our two hypotheses. The first hypothesis concerned whether externalizing behaviors are related to social breadth at each of several consecutive levels of Q (with lower levels of Q representing more short-lived exposure patterns and with higher levels of Q representing greater exposure patterns). The second hypothesis concerned whether externalizing behaviors were related to variations in the pattern of decline in social breadth across levels of Q.

Preliminary Analyses of Social Breadth

Social breadth could be calculated at Q values from 1 through 46 in the present sample (i.e., no child was observed to interact with the same peer on more than 46 days during the semester). However, because of limited variability at the greatest Q levels, data are presented and analyzed for social breadth scores ranging from $Q = 1$ through $Q = 15$ only. Descriptive statistics for social breadth values at $Q = 1$ through $Q = 15$ are presented in Table 1. Because the largest class size for this sample was 20 (range = 17–20), social breadth values could not exceed 19 (i.e., at most, only 19 classmates were available as peer interactants). At $Q = 1$ (e.g., children were observed together at least one time in the semester), social breadth scores ranged from a low of 4 to a high of 19, and the mean of 16.1 indicated that, on average, children interacted with almost every classmate at least once in the semester. Social breadth scores decreased as Q increased,

Table 1

Descriptive Statistics for Social Breadth Scores at Q = 1 Through Q = 15 (N = 97)

Q	Minimum	Maximum	*M*	*SD*	Skewness	Kurtosis
1	4	19	16.1	2.7	−1.4	3.3
2	0	19	13.4	4.2	−0.7	0.3
3	0	19	11.2	5.1	−0.2	−0.8
4	0	19	9.5	5.3	0.2	−1.0
5	0	19	7.8	5.1	0.4	−0.9
6	0	18	6.6	4.8	0.6	−0.6
7	0	17	5.5	4.4	0.8	−0.2
8	0	16	4.5	3.8	0.9	0.0
9	0	16	3.8	3.5	1.2	1.1
10	0	14	3.2	3.1	1.2	1.2
11	0	11	2.8	2.7	1.1	0.9
12	0	11	2.3	2.4	1.3	1.5
13	0	9	2.0	2.2	1.3	1.3
14	0	9	1.7	1.9	1.4	2.0
15	0	8	1.5	1.8	1.5	2.4

Note. Q is operationalized as days (e.g., the value of the function at $Q = 8$ indicates the number of peers observed interacting with a target child on at least 8 days during the semester).

and at $Q = 15$, the mean was 1.5. Note, however, that the maximum score of 8 suggests that some children interacted with as many as eight classmates for at least 15 days. Social breadth scores were normally distributed, and no score exceeded recommended cutoff scores of skewness = 2 or kurtosis = 7 (Curran, West, & Finch, 1996).

Social breadth scores were modestly sensitive to age-related differences among preschoolers. At $Q = 1$, age was not significantly correlated with social breadth. However at $Q = 2$ through $Q = 15$, the correlations between age and social breadth were of modest magnitude and significant at $p < .05$ or better; range was $r(97) = .21, p < .05$ to $r(97) = .31, p < .01$. There were no sex differences in social breadth scores at any level of Q. Because this sample was ethnically diverse, we also tested for ethnicity differences in social breadth, but these were not significant at any level of Q. Children who are more frequently absent from school have fewer opportunities to be observed interacting with peers. Social breadth was significantly correlated with the proportion of time that children were absent from school or unavailable (e.g., in the bathroom or napping), such that children who were absent more frequently had lower social breadth scores at $Q = 1$ through $Q = 15$; range was $r(97) = -.36$ to $r(97) = -.52$, all $ps < .001$. We also calculated the intraclass correlation (ICC) reflecting the effect of classroom differences on social breadth scores at $Q = 1$ through $Q = 15$. ICCs ranged from .40 (at $Q = 15$) to .71 (at $Q = 6$). According to Kreft (1996), ICC scores greater than .10 indicate significant within-group variance on the outcome variable. As such, there was substantial variability in breadth scores across classrooms. Thus, child age, absences/unavailability, and classroom (treated as a random effect) were entered as covariates in subsequent analyses.

Preliminary Analyses of Externalizing Variables

Table 2 provides descriptive statistics for the externalizing-related variables: physical aggression, attention problems, and anger. None of the externalizing-related variables were significantly skewed. Although there were no age or ethnicity differences in any externalizing variable, there were significant sex differences. Boys were significantly more likely to be rated higher on attention problems and anger than were girls, $ts(86, 87) = 3.01$ and 3.18, respectively, $ps < .01$. The effect sizes for these relations were in the large range. Moreover, boys were marginally higher than girls on physical aggression, $t(86) = 1.67$, $p = .10$, with a medium effect size.

Relations Between Externalizing Behaviors and Social Breadth at Each Level of Q

Using SAS (9.1) Proc Mixed, we conducted mixed model regressions to assess the relations between externalizing-related behaviors (physical aggression, attention problems, and anger) and social breadth at $Q = 1$ through $Q = 15$. For each analysis, age and school absences/unavailability were entered as fixed covariates and classroom was entered as a random covariate. Main effects for externalizing behaviors and sex were entered (with separate analyses conducted for physical aggression, attention problems, and anger due to high intercorrelations among the variables); $rs(88)$ ranged from .52 to .80, $ps < .001$. Sex differences were tested by the inclusion of an interaction term between the appropriate externalizing variable and sex. Social breadth at each value of Q was analyzed separately to test for variations in the relations between externalizing behavior and social breadth across increasing levels of Q.

None of the interaction terms with sex reached statistical significance. The interaction terms were then dropped, and the models were reestimated. Table 3 presents the results of the regression analyses for the effects on social breadth attributable to aggressive behavior, attention problems, and anger. The results indicated that the externalizing variables were unrelated to social breadth at low levels of Q. However, at medium to high levels of Q, physical

Table 2
Descriptive Statistics for Externalizing Variables ($n = 97$)

	Total sample						Sex differences		
Variable	Minimum	Maximum	*M*	*SD*	Skewness	Kurtosis	*M* (*SD*) of boys	*M* (*SD*) of girls	Cohen's *d*
Physical Aggression	0.0	36.0	6.4	9.5	1.5	1.3	7.9 (9.5)	4.6 (9.3)	.35
Attention Problems	0.0	14.0	2.5	3.7	1.5	1.3	3.6 (4.1)$_a$	1.3 (2.8)$_a$	.65
Anger	1.6	6.9	3.9	1.4	0.3	−1.0	4.3 (1.3)$_a$	3.4 (1.4)$_a$	.67

Note. Means in the same row that share lowercase subscript letters are significantly different at $p < .01$.

Table 3
Relations Between Externalizing Behaviors and Social Breadth Scores at Q = 1 Through Q = 15

	Physical aggression		Attention problems		Anger	
Q	Est. (*SE*)	*t*	Est. (*SE*)	*t*	Est. (*SE*)	*t*
1	−.00 (0.02)	−0.14	−.01 (0.06)	−0.20	−.01 (0.16)	−0.10
2	−.01 (0.03)	−0.38	−.05 (0.09)	−0.56	−.18 (0.22)	−0.83
3	−.01 (0.03)	−0.43	−.08 (0.09)	−0.88	−.08 (0.23)	−0.35
4	.01 (0.03)	0.22	.01 (0.09)	0.06	−.06 (0.23)	−0.27
5	−.03 (0.03)	−0.88	−.12 (0.08)	−1.43	−.19 (0.21)	−0.90
6	−.03 (0.03)	−1.32	−.15 (0.07)	−2.03*	−.26 (0.18)	−1.40
7	−.05 (0.03)	−1.95	−.18 (0.07)	−2.48*	−.33 (0.18)	−1.81
8	−.04 (0.02)	−1.87	−.16 (0.06)	−2.48*	−.33 (0.16)	−2.00*
9	−.06 (0.02)	−2.49*	−.18 (0.06)	−2.86**	−.38 (0.16)	−2.35*
10	−.06 (0.02)	−2.75**	−.17 (0.06)	−3.00**	−.34 (0.14)	−2.44*
11	−.04 (0.02)	−1.92	−.12 (0.05)	−2.20*	−.21 (0.13)	−1.57
12	−.04 (0.02)	−2.30*	−.12 (0.05)	−2.36*	−.21 (0.12)	−1.74
13	−.04 (0.02)	−2.79*	−.13 (0.04)	−2.94**	−.23 (0.11)	−2.08*
14	−.03 (0.01)	−2.44*	−.10 (0.04)	−2.47*	−.20 (0.10)	−2.07*
15	−.0 3(0.01)	−2.12*	−.07 (0.04)	−1.75	−.23 (0.09)	−2.51*

Note. Age, sex, and school absences were controlled as fixed effects, and classroom was controlled as a random effect.
Note. Degrees of freedom are 1, 78 for analyses with physical aggression and attention problems and 1, 79 for analyses with anger.
$^{*}p < .05$. $^{**}p < .01$.

aggression, attention problems, and anger were negatively related to social breadth. Specifically, for physical aggression, statistically significant effects were found for *Q* = 9 and above (with a marginal effect at $p = .06$ when *Q* = 11). Significant effects were obtained for attention problems at *Q* values greater than 5 (with a marginal effect at $p = .08$ when *Q* = 15), and for anger at *Q* values greater than 7 (with marginal effects of *p*s = .12 and .09 when *Q*s = 11 and 12, respectively). Effect sizes for statistically significant effects were generally in the medium range, ranging from δs = .32 to .42 for physical aggression, from δs = .31 to .45 for attention problems, and from δs = .30 to −.37 for anger. Additionally, the term for school absence was significant at $p < .001$ in each analysis, but the age and sex terms did not reach statistical significance in any analysis. Finally, it is worth noting that classroom level effects were strong, and the average ICC was .66.

Relations Between Externalizing Behaviors and Patterns of Decline in Social Breadth

In the next set of analyses, we considered how patterns of decline in social breadth across levels of exposure (i.e., *Q*) related to externalizing behaviors. Classification techniques, such as cluster analysis, are inherently useful for grouping individuals according to various patterns across variables, and such analytic tools are well suited to graph theory in general and social network analysis in particular (see Arabie & Hubert, 1992, for a review). Typically in developmental science, measures of different constructs taken at one point in time are cluster analyzed to identify subgroups that vary along a number of dimensions. However, in prior research, cluster analysis has also been shown to be a valuable tool for differentiating the profiles of subgroups that differ on a construct across time (Gorman-Smith, Tolan, Loeber, & Henry, 1998; Hirsch & DuBois, 1991). In this case, we adapted this approach by cluster analyzing social breadth scores across increasing *Q* values. It should be noted that the social breadth scores at each level of *Q* were related, such that Q_i equaled Q_{i+1} plus an additional term with random variability. Although cluster analyzing such mathematically dependent variables is novel in the developmental sciences, it does not violate the assumptions of cluster analysis nor is it without precedent in other disciplines, such as biology (Toh & Horimoto, 2002). In the present study, we first used a hierarchical clustering procedure to identify distinct subgroups of children with differing patterns of decline in social breadth scores as exposure increased from *Q* = 1 to *Q* = 15. We followed these analyses with mixed model analyses to assess the relations between externalizing behaviors and variations in patterns of the decline in social breadth.

Cluster analysis of decline in social breadth. We used the average linkage clustering method, which creates homogeneous clusters by minimizing the mean distance between each case and all other cases in the cluster. Squared Euclidean distances were used as the measure of dissimilarity (Milligan, 1996). Prior to cluster analysis, social breadth scores at $Q = 1$ through $Q = 15$ were equated by dividing by their respective ranges. This method is recommended by Milligan and Cooper (1988) and is preferable to standardization of variables because it places variables on the same scale without diminishing differences in variances. Examination of the dendrogram and agglomeration coefficients indicated that three groups could be differentiated. The resulting three-cluster solution was validated by conducting a confirmatory cluster analysis (Fisher & Ransom, 1995). Specifically, we used the *k*-means clustering method (MacQueen, 1967) to conduct a nonhierarchical cluster analysis specifying three groups. Because deriving the clusters' initial centroids randomly may create distortions in the classification scheme in a nonhierarchical cluster analysis (Milligan & Sokol, 1980), we ran the nonhierarchical cluster analysis twice, once without specifying start values and a second time in which we entered start values based on the final cluster centroids that were produced using a third clustering procedure—Ward's (1963) clustering method. The results of both nonhierarchical analyses were identical. We then compared the original cluster solution with the confirmatory cluster solution using the kappa coefficient, $\kappa = .64$, which is in the good range (i.e., $.60 < \kappa < .80$; Landis & Koch, 1977).

The resulting three-cluster solution is depicted in Figure 5. The cluster groups could be differentiated from one another on the basis of mean scores on social breadth, as well as differences in the curve of the decay patterns (e.g., slopes) and are labeled according to mean social breadth scores as *high* (25% of cases), *medium* (14% of cases), and *low* (61% of cases). To validate the results of the cluster analysis, we conducted a multivariate analysis of variance, comparing cluster groups with social breadth scores at *Q* values of 1–15. We obtained a significant multivariate effect, $F(30, 160) = 10.28$, $p < .001$, $\eta^2 = .66$; significant univariate effects were obtained for social breadth scores at each level of *Q*, $Fs(2, 94)$ ranged from 26.23 to 192.74, all $ps < .001$; η^2 estimates ranged from .36 to .80. Post hoc analyses indicated that at $Q = 1$, there were no significant differences in social breadth scores between the high and medium groups, but children in both the high and medium groups interacted with significantly more peers than those in the low group, $ps < .001$. At subsequent levels of *Q* (2 through 15), children in the high group interacted with significantly more peers (i.e., had higher social breadth scores) than children in both the medium and low groups, and children in the medium group interacted with significantly more peers (i.e., had higher social breadth scores) than children in the low group, all $ps < .05$.

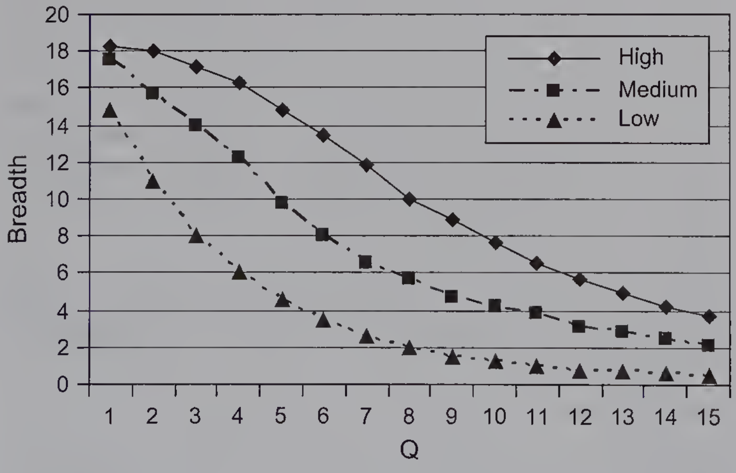


Figure 5. Social breadth scores for three cluster groups at $Q = 1$ through $Q = 15$.
Note. *Q* is operationalized as days (e.g., the value of the function at $Q = 8$ indicates the number of peers observed interacting with a target child on at least 8 days during the semester).

To illustrate the variations in the curves, we calculated the decay rates for each group using an exponential function. We found that an exponential function best fit the overall pattern of the data. Fit indices were R^2s = .98, .99, and .99, for the high, medium, and low groups, respectively. The decay formula for the high group was $y = 25.115e^{-0.1224x}$. This indicates that the estimated intercept for this group was 25.12 (where the intercept reflects the estimated value of social breadth at $Q = 0$), and the estimated decay rate was .12. The decay formula for the medium group was $y = 20.894e^{-0.1558x}$. Thus, the estimated intercept for this group was 20.89 and the estimated decay rate was .16. Finally, the decay formula for the low group was $y = 16.311e^{-0.2485x}$, indicating that the estimated intercept for this group was 16.31 and the estimated decay rate was .25.

Finally, we considered whether the children in the high, medium, and low groups differed on either sex or age. There were no significant sex differences. However, we did obtain a significant effect for age, $F(2, 94) = 4.50$, $p < .05$, $\eta^2 = .09$. Younger children were significantly more likely to be in the low group than in either the medium or high groups, although the effect size for this relation was weak.

Relations between social breadth clusters and externalizing behaviors. We conducted mixed model analyses of variance to estimate the relations between the patterns of decline in social breadth that we obtained from the cluster analyses and externalizing behaviors. Age and school absence/unavailability were controlled as fixed covariates and classroom was controlled as a random effect. Additionally, a main effect term for sex and interaction terms representing the cluster group by sex effect were entered. Externalizing behavior variables served as criterion variables in separate analyses. Significant overall effects were followed with Tukey post hoc tests with the Kramer adjustment for unequal cell sizes (Tukey, 1973).

None of the interaction terms with sex were significant. Thus, the interaction terms were deleted, and the models were reestimated. The mean scores (adjusted for age, sex, school absences, and classroom) on the externalizing-related variables for each cluster group are presented in Table 4. The overall effect of social breadth cluster group on physical aggression approached, but did not reach, statistical significance, $F(2, 77) = 2.86$, $p = .06$. For attention problems, the overall effect of social breadth cluster group was statistically significant, $F(2, 77) = 3.11$, $p = .05$. Children in the medium group had significantly higher scores on attention problems than those in the high group, $t(77) = 2.47$, $p < .05$. A significant overall effect was also obtained for anger, $F(2, 78) = 5.81$, $p < .01$. Again, we obtained a significant difference between children in the medium group and those in the high group, with children in the medium group rated as angrier than children in the high group, $t(78) = 2.49$, $p < .05$. Additionally, we also found that children in the low group were significantly more angry than children in the high group, $t(78) = 3.28$, $p < .01$. Effect sizes were in the small range for attention problems ($\delta = .27$) and in the medium range for anger ($\delta = .36$). The following fixed covariates also reached statistical significance: sex (for models with attention problems and anger) at $p < .01$ and school absence/unavailability (for models with physical aggression and anger) at $p < .05$. The ICCs suggested small to modest classroom effects for models with physical aggression and attention problems (the ICCs were .16 and .32, respectively) but no meaningful classroom effects for the model with anger (the ICC was .00).

Table 4
Mean Scores on Externalizing-Related Behaviors for Social Breadth Cluster Groups

	Low	Medium	High
Physical aggression	7.5 (2.0)	9.5 (3.1)	2.0 (2.9)
Attention problems	2.6 (1.0)	4.2 $(1.3)_a$	1.3 $(1.3)_a$
Anger	4.2 $(0.2)_a$	4.2 $(0.4)_b$	3.1 $(0.3)_{ab}$

Note. Means are adjusted for age, sex, school absences, and classroom. Standard errors are presented in parentheses. Means in the same row that share lowercase subscript letters are significantly different at $p < .05$.

Discussion

In this study, we used a new quantitative procedure, the *Q*-connectivity method, which builds on traditional social network methods by using repeated assessments of children's peer interactions to characterize individual children's social networks. We used this method to assess preschoolers' social breadth as a function of their exposure to peers over time, and we demonstrated relations between social breadth and three indices of externalizing behaviors—physical aggression, attention problems, and anger.

Externalizing Behavior and Social Breadth

We examined the relations between externalizing behavior and the drop-off in social breadth that occurs in relation to increasing exposure to peer interaction (e.g., value of *Q*). We considered drop-off in two ways: *decreases in social breadth* scores across *Q* values and *variations in patterns of decline* across *Q* values. Analyses of the relations between externalizing behaviors and social breadth scores across *Q* values demonstrated that several aspects of externalizing behavior—physical aggression, attention problems, and anger—were negatively related to social breadth. However, these relations only emerged at moderate to high levels of *Q*—that is, moderate to high frequency of exposure to the same peers. These relations were not evident at lower levels of exposure. The overall pattern of findings was robust along two important dimensions. First, the findings were consistent for behavioral (physical aggression), attentional (attention problems), and emotional (anger) indices of externalizing behavior, suggesting consistency across constructs. Second, the consistent lack of effects at lower levels of *Q* combined with the significant findings at moderate to higher levels of *Q* (with the exception of a few instances in which the effects approached, but fell just short of reaching, statistical significance; these were $Q = 11$ for physical aggression, $Q = 15$ for attention problems, and $Qs = 11$ and 12 for anger) suggests consistency in the patterns of relations as exposure to peers increases. Moreover, the obtained effect sizes in the medium range suggest that these are not trivial findings.

Analyses of patterns of decline in social breadth suggested that three groups of children could be differentiated, with group trajectories defined according to both intercept and curve. Children who exhibited a pattern of social breadth characterized by initially high social breadth scores (i.e., a high intercept) and a modestly declining curve tended to have relatively high scores on attention problems and anger, although the relation with physical aggression did not quite reach statistical significance. In addition, children who exhibited a pattern characterized by initially low social breadth scores and a steeply declining curve tended to have relatively high scores on anger but not on physical aggression or attention problems. In contrast, those children who exhibited a pattern of social breadth characterized by initially high scores and a slowly declining curve were unlikely to exhibit physical aggression, attention problems, or anger.

Taken together, these findings suggest that externalizing preschoolers, like their classmates, may play with most peers in the class on one or a few occasions. However, their exposure to peers moderately to rapidly became selective as they began to interact with a relatively small number of peers. Of particular interest is the question of what peer processes underlie the drop-off in peer exposure for externalizing children, keeping in mind that these processes may vary depending on the type of peer with which an externalizing child is interacting. Either of two social processes may be at play here. On the one hand, if externalizing children prefer playing with peers with similar levels of aggression, they may have few peers to choose from in a class. Furthermore, playing with peers who are not similar on this dimension of behavior may not be enjoyable; thus, exposure to nonaggressive peers may drop off quickly. Thus, on the one hand, this effect may be driven by externalizing children themselves. On the other hand, such children may be actively disliked by the larger peer group (Snyder et al., 1997). Thus, they might have fewer opportunities to engage with mainstream peers as their classmates become sensitized to their negative behaviors and actively reject them (Olson, 1992).

It is likely that both processes may contribute to the drop-off that we found in this study. However, these processes may be differentially weighted or they may operate differently depending on the type of peer—internalizing versus socially competent versus externalizing. For example, one recent study mapped the dynamic nature of externalizing children's tendencies to interact with externalizing, socially competent, and internalizing peer partners (Martin et al., 2005). Findings from this study demonstrated that interactions with peers who were themselves externalizing were attractor states for young externalizing boys and girls (providing evidence for behavioral homophily) and, to a somewhat lesser degree, interactions with socially competent peers were also attractor states. However, externalizing children rarely interacted with internalizing peers. Perhaps the interactions with internalizing peers are the most ephemeral for young externalizing children (dropping off quickly), whereas the interactions with other externalizing children are most likely to be repeated. Presumably, then, internalizing peers are those who only show up at low levels of exposure (e.g., $Q = 1$ or 2). Socially competent peer partners might show up at moderate levels of exposure (e.g., moderate Q graphs), and those peers in the highest Q graphs are likely to be the externalizing ones. Thus, homophily and selectivity might underlie social breadth patterns in relation to peers who are also externalizing, but rejection might explain the drop-off in social breadth for peers who are socially competent. Future research that tests these hypotheses with a fine-grained analysis of the behaviors that are emitted by both externalizing children and their peers during interactions could shed more light on this issue. Understanding these issues can provide insight into the social trajectories on which young externalizing children find themselves; this is necessary to attempts to intervene in this behavioral pathway.

Developmental–Contextual Perspective on Social Breadth

The primary research findings should be interpreted in light of the developmental–contextual effects that were (and were not) obtained. First, there were neither main effects nor interaction effects of sex on social breadth, indicating that boys and girls had similar social structures, at least with respect to the breadth of their peer relationships and at least with the sample of young children that we used. There has been some question about the extent to which there are sex differences in the size of children's peer networks. Studies on kindergarten-age and older children have suggested that boys have larger social networks than do girls (Ruble et al., 2006). Moreover, Waldrop and Halverson (1975) found that 7-year-old boys had more extensive peer relationships than did girls. Among younger children, however, the findings are mixed, with some studies suggesting that there are sex differences in the size of social networks and others finding no differences (Ruble et al., 2006). This may reflect a developmental process; thus, we speculate that sex differences in social breadth would be apparent among older children.

Moreover, the lack of sex effect in the present study should be seen as contributing to recent efforts to explore sex differences in externalizing behaviors and the social correlates of externalizing behavior. Consistent with the larger body of research on physical aggression, in the present study, boys were rated as more externalizing than girls, that is, boys were rated by teachers as significantly more angry and as displaying significantly more attention problems, although the sex difference in physical aggression, favoring boys, approached, but did not reach, statistical significance. This finding is generally robust, as opposed to findings regarding sex differences in relational forms of aggression, which tend to be mixed, even among young children (Crick, Casas, & Mosher, 1997; Hart, Nelson, Robinson, Olsen, & McNeilly-Choque, 1998; Underwood, Galen, & Paquette, 2001). Given the inherently social nature of relational aggression, it would be useful to replicate this research using measures of relational aggression, as the patterns may vary. Moreover, this study speaks to a small, but important, body of literature that has considered whether there are sex differences in the quality of peer relations for aggressive children. Most of the extant research in this area has focused on peer rejection, and the findings have been mixed, with some studies showing sex differences in the extent to which externalizing children are rejected and others showing no sex differences (Zimmer-Gembeck et al., 2005). The lack of significant sex effects in the present study may indicate that the general structures of peer interactions for young externalizing boys and girls are quite similar despite any variations in involvement in physical forms of aggression. More research on this issue is clearly needed.

The present findings also support the idea that social breadth may vary with age. In the 3- to 5-year-old children in the present study, we found age-related differences in social breadth. This developmental period is an important one from a social interactional perspective. At this age, children are making the shift from playing alone or alongside peers to playing with peers (Rubin & Coplan, 1998). It was no surprise, then, that younger children in our sample had lower social breadth than did older children and that they demonstrated the steepest drop-off in social breadth. This may typify a normative social developmental process in early childhood that is marked by the movement from solitary play behaviors toward increasing engagement in social play with age and social experience. Recall that social breadth was measured in the fall semester. It is possible that assessments of social breadth in the spring semester of preschool (after children had several months of experience in peer group settings) would identify fewer children in the low group. A question for future research is whether social breadth continues to show age-related variations as children get older; perhaps age differences would be seen as children reach important developmental transitions, such as the transition from elementary to middle school or the transition from engaging in same-sex peer groups to greater involvement with members of the other sex that occurs in adolescence.

From a contextual perspective, we obtained relatively strong classroom effects, suggesting that aspects of the classroom environment contribute to children's socialization patterns. Classroom characteristics may affect social breadth in numerous ways. Perhaps the aggregation of particular types of peers (e.g., classrooms with larger numbers of externalizing children) creates variations in socialization patterns. As demonstrated by Stormshak et al. (1999), social preference scores in first-grade classrooms depended upon the extent to which classmates were aggressive, with aggressive children having higher social preference scores in aggressive classrooms and lower social preference scores in nonaggressive classrooms. Additionally, we have previously reported classroom level effects on patterns of peer victimization in preschool classrooms using a different sample of young children (Hanish, Ryan, Martin, & Fabes, 2005). Alternatively, teachers' behaviors and child management styles may create different patterns of socialization from one classroom to another. For instance, the extent to which teachers take an active role in structuring children's peer relationships may influence socialization patterns (Howes, Hamilton, & Matheson, 1994). We did not have a large enough sample of classrooms to adequately test for classroom effects in this study. Thus, future research that dissects the role of classrooms is needed to better understand this contextual contributor to social breadth.

Future Directions for the Use of the Q-*Connectivity Method*

In the present study, we did not directly consider issues related to peer contagion processes, although the findings regarding social breadth set the stage for a process-oriented analysis. Recent evidence suggests that, even in early childhood, the opportunity exists for both selection and influence processes to operate with regard to the spread of aggressive behaviors (Hanish, Martin, et al., 2005). Little is known about the extent to which weak ties with peers (e.g., brief encounters, acquaintanceships) as well as strong ties with peers (e.g., mutual friendships) are determined by peer selection processes and about the extent to which each enables peer socialization along externalizing dimensions. However, one recent study with adolescents suggests that there is value in considering the extent to which weak social ties influence decisions to engage in delinquent behaviors (Bauman, Faris, Ennett, Hussong, & Foshee, 2007). Building on this idea, future studies that explore externalizing children's social breadth in more depth, for instance, with regard to the valence of the interactions (e.g., Do the patterns of social breadth depend on the extent to which peer interactions are negative or positive?) or the characteristics of peers (e.g., Is social breadth moderated by peers' characteristics?) are possible using the *Q*-connectivity method and would likely be fruitful.

An additional avenue for future research has to do with the conceptual and analytic approach to the time dimension. One of the most interesting and important features of the *Q*-connectivity method is that it incorporates time in a direct way. There are numerous ways to reflect time with this method. Some of these include conducting frequency counts over some period of time and calculating trajectories over *Q*, as we did here. An example of another use of the time component involves creating moving windows that can assess variations in peer structures across time, which would allow a more fine-grained assessment of the time between contacts with various peers. Although this issue was not central to the theme and focus of the present article, it is of great interest to a number of important developmental questions. For instance, one might hypothesize that frequent contacts with a peer that occur at the beginning of the school year but that diminish over time might have a different developmental meaning than an equal number of contacts with a peer that are spaced out over the academic semester or year. Future research that builds on the present study by considering time in alternative ways would lend insights into the development of peer relationships and contribute to the small, but growing, compendium of dynamic methods for studying peer relations (e.g., state space grids, Granic & Patterson, 2006; Simulation Investigation for Empirical Network, Snijders, 2005).

One unique feature of this study was the use of the *Q*-connectivity method to map each individual child's social landscape and how this changed across differing levels of exposure to peers. We focused only on one parameter of *Q*-connectivity, namely social breadth. Although it is beyond the scope of the current paper, the *Q*-connectivity method also has the potential to offer a host of informative measures of interactional patterns; connectivity, density, maximum value of *Q*, dyadicity, and a variety of other social network measures across *Q* are examples of just a few measures that can be calculated using this method. *Q*-connectivity analyses also may be undertaken to reveal latent structures in social interactional patterns. Additionally, the use of graphics as a data visualization tool for examining social interactional patterns over time may make a powerful contribution and may open up new lines of systemic research (Legrand, 2002). The introduction of the *Q*-connectivity method allows for an expanded ability to assess peer interactions and thus has the potential to inform researchers about peer socialization processes. Thus, the *Q*-connectivity method has relevance to issues such as understanding the formation, evolution, and dissolution of peer relationships, peer selection effects, and peer influence processes.

References

Achenbach, T., & Rescorla, L. A. (2000). *Manual for the ASEBA preschool forms and profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, and Families.

Arabie, P., & Hubert, L. J. (1992). Combinatorial data analysis. *Annual Review of Psychology, 43*, 169–203.

Asakawa, K., & Csikszentmihalyi, M. (2000). Feelings of connectedness and internalization of values in Asian American adolescents. *Journal of Youth and Adolescence, 29*, 121–145.

Barcelo, H., & Laubenbacher, R. (2005). Perspectives on a homotopy theory and its applications. *Journal of Discrete Mathematics, 298*, 39–61.

Bauman, K. E., Faris, R., Ennett, S. T., Hussong, A., & Foshee, V. A. (2007). Adding valued data to social network measures: Does it add to associations with adolescent substance use? *Social Networks, 29*, 1–10.

Birnbaum, G. E., Reis, H. T., Mikulincer, M., Gillath, O., & Orpaz, A. (2006). When sex is more than just sex: Attachment orientations, sexual experience, and relationship quality. *Journal of Personality and Social Psychology, 91*, 929–943.

Cairns, R. B., Cairns, B. D., Neckerman, H. J., Gest, S. D., & Gariepy, J. L. (1988). Social networks and aggressive behavior: Peer support or peer rejection? *Developmental Psychology, 24*, 815–823.

Cairns, R. B., Leung, M., Buchanan, L., & Cairns, B. D. (1995). Friendships and social networks in childhood and adolescence: Fluidity, reliability, and interrelations. *Child Development, 66*, 1330–1345.

Challman, R. C. (1932). Factors influencing friendships among preschool children. *Child Development, 3*, 146–158.

Crick, N. R., Casas, J. F., & Mosher, M. (1997). Relational and overt aggression in preschool. *Developmental Psychology, 33*, 579–588.

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods, 1*, 16–29.

Eagle, N., & Pentland, A. (2006). Reality mining: Sensing complex social systems. *Personal and Ubiquitous Computing, 10*, 255–268.

Estell, D. B., Cairns, R. B., Farmer, T. W., & Cairns, B. D. (2002). Aggression in inner-city early elementary classrooms: Individual and peer-group configurations. *Merrill-Palmer Quarterly, 48*, 52–76.

Fabes, R. A., Shepard, S. A., Guthrie, I. K., & Martin, C. L. (1997). Roles of temperamental arousal and gender-segregated play in young children's social adjustment. *Developmental Psychology, 33*, 693–702.

Fisher, L., & Ransom, D. C. (1995). An empirically derived typology of families: l. relationships with adult health. *Family Process, 34*, 161–182.

Goldsmith, H. H., & Rothbart, M. K. (1991). Contemporary instruments for assessing early temperament by questionnaire and in the laboratory. ln J. Strelau & A. Angleitner (Eds.), *Explorations in temperament: International perspectives on theory and measurement. Perspectives on individual differences* (pp. 249–272). New York: Plenum Press.

Gorman-Smith, D., Tolan, P. H., Loeber, R., & Henry, D. B. (1998). Relation of family problems to patterns of delinquent involvement among urban youth. *Journal of Abnormal Child Psychology, 26*, 319–333.

Granic, I., & Patterson, G. R. (2006). Toward a comprehensive model of antisocial development: A dynamic systems approach. *Psychological Review, 113*, 101–131.

Hanish, L. D., Barcelo, H., Martin, C. L., Fabes, R. A., Holmwall, J., & Palermo, F. (2007). Moving beyond the dyad: Using the Q-connectivity method to examine preschoolers' peer group interactions. In R. W. Larson & L. A. Jensen (Series Eds.) & P. C. Rodkin & L. D. Hanish (Vol. Eds.), *New directions for child and adolescent development: Peer social networks* (Vol. 118, pp. 9–24). San Francisco: Jossey-Bass.

Hanish, L. D., Martin, C. L., Fabes, R. A., Leonard, S., & Herzog, M. (2005). Exposure to externalizing peers in early childhood: Homophily and peer contagion processes. *Journal of Abnormal Child Psychology, 33*, 267–281.

Hanish, L. D., Ryan, P., Martin, C. L., & Fabes, R. A. (2005). The social context of young children's peer victimization. *Social Development, 14*, 2–19.

Hart, C. H., Nelson, D. A:, Robinson, C. C., Olsen, S. F., & McNeilly-Choque, M. K. (1998). Overt and relational aggression in Russian nursery-school-age children: Parenting style and marital linkages. *Developmental Psychology, 34*, 687–697.

Hirsch, B. J., & DuBois, D. L. (1991). Self-esteem in early adolescence: The identification and prediction of contrasting longitudinal trajectories. *Journal of Youth and Adolescence, 20*, 53–72.

Howes, C., Hamilton, C. E., & Matheson, C. C. (1994). Children's relationships with peers: Differential associations with aspects of the teacher-child relationship. *Child Development, 65*, 253–263.

Kreft, I. G. G. (1996). Are multilevel techniques necessary? An overview, including simulation studies. Retrieved January 2003 from http://www.stat.ucla.edu/~kreft/-quarterly/quarterly.html

Landis, J. R., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159–174.

Le, B., Hat, N. C., & Beal, D. J. (2006). Pocket-sized psychology studies: Exploring daily diary software for palm pilots. *Behavior Research Methods, 38*, 325–332.

Legrand, J. (2002, October). How far can Q-analysis go into social systems understanding? ln *Proceedings of the Fifth European Systems Science Congress* (pp. 1–10). Crete.

MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics, 1*, 281–297.

Magnusson, D. (1998). The logic and implications of a person-oriented approach. In R. B. Cairns, L. R. Berfman, & J. Kagan (Eds.), *Methods and models for studying the individual* (pp. 33–64). Thousand Oaks, CA: Sage.

Martin, C. L., & Fabes, R. A. (2001). The stability and consequences of young children's same-sex peer interactions. *Developmental Psychology, 37*, 431–446.

Martin, C. L., Fabes, R. A., Hanish, L. D., & Hollenstein, T. (2005). Social dynamics in the preschool. *Developmental Review, 25*, 299–327.

McHale, S. M., Crouter, A. C., & Bartko, T. (1992). Traditional and egalitarian patterns of parental involvement: Antecedents, consequences, and temporal rhythms. In D. Featherman, R. Lerner, & M. Perlmutter (Eds.), *Life-span development and behavior* (Vol. 11, pp. 49–81). Hillsdale, NJ: Lawrence Erlbaum.

Milligan, G. W. (1996). Clustering validation: Results and implications for applied analysis. In P. Arabie, L. J. Jubert, & G. De Soto (Eds.), *Clustering and classification* (pp. 341–375). River Edge, NJ: World Scientific.

Milligan, G. W., & Cooper, M. C. (1988). A study of standardization of variables in cluster analysis. *Journal of Classification, 5*, 182–204.

Milligan, G. W., & Sokol, L. M. (1980). A two-stage clustering algorithm with robust recovery characteristics. *Educational and Psychological Measurement, 40*, 755–759.

Moody, J., McFarland, D., & Bender-deMoll, S. (2005). Dynamic network visualization. *American Journal of Sociology, 110*, 1206–1241.

Olson, S. L. (1992). Development of conduct problems and peer rejection in preschool children: A social systems analysis. *Journal of Abnormal Child Psychology, 20*, 327–350.

Richards, W. D., & Rice, R. E. (1981). The NEGOPY network analysis program. *Social Networks, 3*, 215–223.

Rodkin, P. C., Farmer, T. W., Pearl, R., & Van Acker, R. (2000). Heterogeneity of popular boys: Antisocial and prosocial configurations. *Developmental Psychology, 36*, 14–24.

Rothbart, M. K., Ahadi, S. A., & Hershey, K. L. (1994). Temperament and social behavior in childhood. *Merrill-Palmer Quarterly, 40*, 21–39.

Rubin, K. H., Bukowski, W., & Parker, J. G. (1998). Peer interactions, relationships, and groups. In W. Damon & N. Eisenberg Lerner (Eds.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (pp. 619–700). New York: Wiley.

Rubin, K. H., & Coplan, R. J. (1998). Social and nonsocial play in childhood: An individual differences perspective. In O. N. Saracho & B. Spodek (Eds.), *Multiple perspectives on play in early childhood education* (pp. 144–170). Albany, NY: State University of New York Press.

Ruble, D. N., Martin, C. L., & Berenbaum, S. A. (2006). Gender development. In N. Eisenberg, W. Damon, & R. Lerner (Eds.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (pp. 858–932). New York: Wiley.

Snijders T. A. B. (2005). Models for longitudinal network data. In P. J. Carrington, J. Scott, & S. Wasserman (Eds.), *Models and methods in social network analysis* (pp. 215–247). Cambridge, UK: Cambridge University Press.

Snyder, J., Horsch, E., & Childs, J. (1997). Peer relationships of young children: Affiliative choices and the shaping of aggressive behavior. *Journal of Clinical Child Psychology, 26*, 145–156.

Stein, C. R., Devore, R. B., & Wojcik, B. E. (2005). Calculation of the kappa statistic for inter-rater reliability: The case where raters can select multiple responses from a large number of categories, SUGI30, Paper 180. In *Proceedings of the Thirtieth Annual SAS® Users Group International Conference* (pp. 1–8). Cary, NC: SAS Institute Inc.

Stormshak, E. A., Bierman, K. L., Bruschi, C., Dodge, K. A., Coie, J. D., & The Conduct Problems Prevention Research Group (1999). The relation between behavior problems and peer preference in different classroom contexts. *Child Development, 70*, 169–182.

Strayer, F. F., & Santos, A. J. (1996). Affiliative structures in preschool peer groups. *Social Development, 5*, 117–130.

Toh, H., & Horimoto, K. (2002). Inference of a genetic network by a combined approach of cluster analysis and graphical Gaussian modeling. *Bioinformatics, 18*, 287–297.

Tukey, J. W. (1973). *The problem of multiple comparisons*. Princeton, NJ: Princeton University.

Underwood, M. K., Galen, B. R., & Paquette, J. A. (2001). Top ten challenges for understanding gender and aggression in children: Why can't we all just get along? *Social Development, 10*, 248–266.

van den Oord, E. J. C. G., Rispens, J., Goudena, P. P., & Vermande, M. (2000). Some developmental implications of structural aspects of preschoolers' relations with classmates. *Journal of Applied Developmental Psychology, 21*, 619–669.

Vespo, J. E., Kerns, K. A., & O'Connor, E. M. (1996). Social organization in preschool classrooms: Social sets as social units. *Journal of Social and Personal Relationships, 13*, 41–55.

Waldrop, M. F., & Halverson, C. F. (1975). Intensive and extensive peer behavior: Longitudinal and cross-sectional analyses. *Child Development, 46*, 19–26.

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association, 58*, 236–244.

Zimmer-Gembeck, M. J., Geiger, T. C., & Crick, N. R. (2005). Relational and physical aggression, prosocial behavior, and peer relations: Gender moderation and bidirectional associations. *Journal of Early Adolescence, 25*, 421–452.

Child Development, July/August 2008, Volume 79, Number 4, Pages 1137 – 1151

The Effects of Background Television on the Toy Play Behavior of Very Young Children

Marie Evans Schmidt, Tiffany A. Pempek, Heather L. Kirkorian, Anne Frankenfield Lund, and Daniel R. Anderson
University of Massachusetts at Amherst

This experiment tests the hypothesis that background, adult television is a disruptive influence on very young children's behavior. Fifty 12-, 24-, and 36-month-olds played with a variety of toys for 1 hr. For half of the hour, a game show played in the background on a monaural TV set. During the other half hour, the TV was off. The children looked at the TV for only a few seconds at a time and less than once per minute. Nevertheless, background TV significantly reduced toy play episode length as well as focused attention during play. Thus, background television disrupts very young children's play behavior even when they pay little overt attention to it. These findings have implications for subsequent cognitive development.

Consider a 2-year-old child playing with toys in the living room. The television is playing a program designed for adult viewers. Mother is in the kitchen making dinner while listening to the TV. The child intermittently looks at the television for brief periods of time. Is television having an effect on the child? In this article, we test the hypothesis that background television is a disruptive influence on very young children's behavior, specifically object play.

Prior research indicates that young children pay little attention to television that they cannot comprehend (e.g., Anderson, Lorch, Field, & Sanders, 1981) and little attention to adult TV programs in general (Schmitt, Anderson, & Collins, 1999). Thus, an adult television program is essentially in the background as the child plays with toys or engages in social interactions. Here, we adopt the term *background television* to refer to adult content that is largely incomprehensible to a very young child and to which they ordinarily pay little cumulative attention (see Anderson & Evans, 2001). Nevertheless, as a dynamically varying audiovisual distraction, background television may interfere with the ability of very young children to sustain an activity in a focused and organized manner.

The American Academy of Pediatrics, Committee on Public Education (1999) recommends no screen media exposure for children aged 2 years and younger. Nevertheless, nearly 75% of parents of very young children say that television is on "about half of the time" or more, even if no one is watching (Rideout & Hamel, 2006; Rideout, Vandewater, & Wartella, 2003). There is substantial opportunity in these homes for background television to have a chronic disruptive impact on very young children's behavior.

It should be emphasized that the present study is concerned with the impact of adult programming that is in the background from the perspective of the child. Very young children certainly spend time watching programs made for them. Studies that did not specifically assess background TV have found that very young children watch from 1 to 3 hr a day, on average (Anderson, Lorch, Collins, Field, & Nathan, 1986; Carew, 1980; Clarke-Stewart, 1973; Gottfried, 1984; Lemish, 1987; Rideout & Hamel, 2006; Rideout et al., 2003; Zimmerman, Christakis, & Meltzoff, 2007), although there are large percentages of children whose parents claim they do not watch any TV (Certain & Kahn, 2002; Christakis, Zimmerman, DiGiuseppe, & McCarty, 2004; Rideout et al., 2003; Woodard, 2000).

Marie Evans Schmidt is now at the Center on Media and Child Health, Children's Hospital, Boston, MA.

Tiffany A. Pempek is now at Georgetown University in Washington, DC.

This research is based in part on a University of Massachusetts doctoral dissertation by Marie Evans Schmidt. Aspects of this research were presented at the meetings of the Society for Research in Child Development and at the meetings of the International Conference on Infant Studies. This research was supported by grants from the National Science Foundation (BCS-0111811 and BCS-0519197). Findings and opinions expressed in this article do not reflect endorsement by the National Science Foundation. We wish to acknowledge the videotape coders for this project: Amy Fuller, Sean Kennedy, Lauren Murphy, Rae Gallagher, Carolynn Laurenza, David Massuda, Rebecca Edwards, Jill Rosenbaum, Kate Decker, Emily Duclos, Erica Gentuso, Kristian Lundberg, Alexis Lauricella, and Angie Naniot. We also wish to acknowledge Aline Sayer and Arnold Well for statistical consultation.

Correspondence concerning this article should be addressed to Daniel R. Anderson, Department of Psychology, University of Massachusetts at Amherst, Amherst, MA 01003. Electronic mail may be sent to anderson@psych.umass.edu.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7904-0021

In the only study to specifically consider background TV, Pierroutsakos, Hanna, Self, Lewis, and Brewer (2004) asked 100 upper-middle-class parents to keep diaries of their 2.5- to 24-month-old infants' TV exposure. The children were exposed to an average of 120 min of TV each day; 49% of that exposure was to adult and preteen programming. It is likely that a broad survey would reveal that a large percentage of very young children are exposed to much more background TV than found by Pierroutsakos et al. insofar as in-home automated monitoring of TV use by Nielsen Media Research indicates that the average American home has a TV set in use more than 8 hr a day (Gertner, 2005).

Effects of Television on Very Young Children

There is very little experimental research on the effects of television on infants and toddlers and no direct research on the effects of background exposure. The studies that have examined long-term effects of early television exposure are correlational and do not account for content of the programs, thus combining exposure to age-appropriate and background television. Television exposure for children aged 30 months and younger is associated with poorer cognitive and language development (Carew, 1980; Gottfried, 1984; Nelson, 1973; Wachs, 1985; Wachs & Gandour, 1983; Wachs & Gruen, 1982; Zimmerman & Christakis, 2005). One study reported that language development depended on which child-directed programs were regularly watched, with some programs positively associated with language development and other programs negatively associated (Linebarger & Walker, 2005).

Christakis et al. (2004) reanalyzed data from the 1980s National Longitudinal Survey of Youth to determine if there was a relationship between early TV viewing and later attention disorders. Parents' estimates of their children's TV viewing at 18 and 42 months were positively associated with parental reports of attention disorder symptoms at age 7 after a variety of covariates were statistically controlled. An analogous study in Denmark, however, failed to find a significant relationship, although the trend was in the same direction (Obel et al., 2004). Television viewing by older preschoolers does not predict attention disorder symptoms (Stevens & Mulsow, 2006).

Several studies have revealed that the level of ambient background noise in the home and household chaos are negatively related to cognitive development in the first 5 years of life (Gottfried, 1984; Wachs, 1985, 1986; Wachs & Gandour, 1983). Because television is a major contributor to the level of ambient noise in these studies, it may play a disruptive role, as hypothesized. Nevertheless, the studies cited above that found negative outcomes associated with television exposure do not allow causal relationships with television to be conclusively inferred.

Background Television and Toy Play

Based on videotapes of family TV viewing recorded in homes, Schmitt, Woolf, and Anderson (2003) reported that 2-year-olds played with toys 32% of their time spent with television. The present study, by experimentally manipulating the presence of background television, is designed to determine whether television is a proximal disruptive influence on the organization of very young children's behavior, specifically toy play. Television may have such an influence by initiating repeated orienting reactions to the visual and auditory changes that frequently occur on commercial television. Television may also be viewed as competing for cognitive resources necessary to instantiate and execute play schemes (cf. Armstrong & Greenberg, 1990). In addition, as a source of environmental noise, background television may have a general disruptive effect that has been observed on children in noisy environments (Hygge, Evans, & Bullinger, 2002; Wachs, 1986).

Virtually every theory of child development hypothesizes that play is related to healthy cognitive and social development (e.g., Piaget, 1962). During play, children refine motor skills, explore the physical properties of objects, learn principles of cause and effect, and engage in means – ends problem solving (Power, 2000). Pretend play, in particular, helps children develop representational abilities as well as experiment with social roles (Bretherton, 1984; Power, 2000). If background television disrupts young children's play, chronic exposure to background television may at least partially be a cause of the negative relationship between early exposure to television and cognitive development.

In this research, we examine play maturity, play episode length, and focused attention during play. With respect to play maturity, we use a scale based on research by Belsky and Most (1981). The logic of the scale extends from numerous studies that reveal a developmental sequence of toy play behaviors. As development proceeds from early infancy, visually guided manipulation and object-appropriate behaviors increase and become more symbolic and complex (Belsky & Most, 1981; McCall, 1974).

Play episode length refers to the time that elapses from the point at which the child comes into contact with a toy until the child ceases active play with that toy (Choi & Anderson, 1991). As children mature, play episode length increases (Malone, 1997; Ruff & Lawson, 1990), probably reflecting more sequentially complex play schemes as well as an ability to sustain attention. Short play episode length is a predictor of attention deficit disorders and other developmental problems (e.g., Faden & Graubard, 2000; Handen, McAuliffe, Janosky, Feldman, & Breaux, 1998; Malone, 1997).

The intensity of attentional engagement may vary during play episodes. The term *focused attention* has been used to designate activity during which "attention is directed more or less exclusively to one target or task and not divided or shared between targets or tasks" (Ruff & Rothbart, 1996, p. 111). As motivation and attention are linked (Derryberry & Tucker, 1990), focused attention usually results when a target activity is highly motivating and is made possible by inhibiting responses to all but the object or task (Ruff, Capozzoli, & Saltarelli, 1996).

Infants and toddlers are capable of focused attention, although the underlying processes supporting attention change substantially during the first 3 years. Ruff and Rothbart (1996) describe the development of two different attentional systems in very young children. The orienting/investigative system, dominant for the 1st year of life, is primarily responsive to novelty, meaning attention in young infants is governed by the salient perceptual features of environmental stimuli. Ruff, Saltarelli, Capozzoli, and Dubiner (1992) found that focused attention in 5- to 12-month-old infants was greatest for novel objects. After 2–3 min of exposure to the same object, however, episodes of focused attention were terminated. Focused attention episodes in infancy are thus usually short, occurring intermittently with other exploratory actions (e.g., banging or mouthing).

The second system of attention, that of higher level control, appears at the end of the 1st year as the prefrontal cortex matures. Ruff and Rothbart (1996) hypothesize that the second system facilitates or inhibits the orienting/investigative system as children develop the cognitive ability to formulate plans or goals. These plans can prevent habituation by facilitating sustained attention after objects are no longer novel. As children develop cognitively and acquire language, their goals can become more sophisticated leading to longer episodes of focused attention, particularly during play. Consistent with this theory, focused attention during play increases with age (Ruff & Capozzoli, 2003; Ruff & Lawson, 1990). With respect to the present research, because very young children have poorly developed higher level control, it is expected that they have a poor ability to filter irrelevant stimuli as they engage in object play.

In light of research on attention development, background TV should disrupt sustained and focused attention during play. This hypothesized disruption occurs because TV repeatedly elicits orienting responses (ORs), thereby drawing attention away from the play. Physiologically, the OR is characterized by heart rate decrease, dilation of blood vessels that lead to the brain, and constriction of blood vessels that lead to the major muscle groups, along with visual orientation in the direction of the stimulation that elicited the OR. The cardiovascular changes are generally thought to reflect shifts in blood supply to the brain in service of alert investigation of novel stimuli. Consistent with previous OR research, Reeves, Thorson, and Schleuder (1986) found decreases in alpha waves that were time locked to formal features (e.g., cuts, sudden camera changes, movement, sound effects) on electroencephalographic (EEG) recordings of adults watching television. Richards and his colleagues (Richards & Casey, 1992; Richards & Cronise, 2000; Richards & Gibson, 1997) also found evidence of the OR produced by television in studies of heart rate in infants and toddlers.

A child who orients to a TV screen may suspend play in order to attend to television. Because there is a visual or auditory change on TV approximately every 6 s (Schmitt et al., 1999), background television could disrupt children's play many times over the course of a play session. A few studies suggest that brief interruptions can terminate young children's ongoing activities. DiLalla and Watson (1988) observed 2.5- to 6.5-year-old children's reactions to three experimenter-initiated interruptions to an ongoing fantasy play sequence. Older children were better able to respond to interruptions without disrupting their subsequent play. The 3-year-olds were unable, however, to return to the play episode without experimenter prompts. Furthermore, if the play resumed, it was less intense. To the extent that it has a disruptive influence, background television may prevent young children from returning to play at the same level of cognitive sophistication, if at all.

It is possible that television does not have these disruptive effects insofar as Ruff and Capozzoli (2003) found that an intermittent brief distractor did not reduce focused attention during toy play in very young children. This may have happened because the children became progressively less distractible over the 12-min course of the play session, suggesting that they habituated to the distractor. However, television, as a complex, continuously varying distractor, may have more substantial effects and may prove to be resistant to habituation.

In the present research, 12-, 24-, and 36-month-old children were observed playing with toys for 1 hr. Children in this age range are just beginning complex and symbolic play and have poorly developed control over focused and sustained attention (Ruff & Rothbart, 1996). For half the time, a TV set played the television program *Jeopardy!* with commercials; for the other half, the TV set was off. The children were videotaped, and observers subsequently coded the tapes for looking at television, length of play episodes, length of focused attention during play episodes, and maturity of play. We hypothesized that background television would disrupt the children's toy play, shortening the length of play episodes and reducing the percent of play that was focused. We also hypothesized that background television would reduce the maturity of children's play.

Unlike prior research on television's impact, we focused specifically on background TV insofar as *Jeopardy!*, a game show directed at adults, should be nearly incomprehensible to such young children. Because many children are exposed to background television in the home and some studies suggest such exposure may negatively influence cognitive and language development, this study is a first step toward understanding one mechanism by which television may affect development.

Method

Participants

Fifty children (1 Black, 2 Hispanic, and the remainder White), 17 (9 girls) 12-month-olds, 16 (8 girls) 24-month-olds, and 17 (9 girls) 36-month-olds, were randomly assigned to either the "TV first" or the "TV second" order with the constraint that half were assigned to each. All children were within 1 month of their first, second, or third birthdays. Seventy-eight percent of children had a participating caregiver whose educational level could be classified as "some college" or higher, with 60% having completed college and 12% having attained a graduate degree.

In general, these ages (12, 24, or 36 months) were chosen because a minimal amount of research has been conducted on the impact of television on very young children. More specifically, the 36-month-old group was chosen because this age overlaps with the youngest ages used in some previous work on children and television. The two younger groups were chosen because they were young enough to ensure that the program used in this study would, in fact, be background television for them insofar as children pay little attention to content that they do not understand (Anderson et al., 1981).

Potential participants were identified through state birth records. Children were recruited by means of a letter describing the study to parents and a follow-up telephone call requesting participation. The only selection criterion was that children had no parent-identified visual or hearing impairments. No participants were excluded based on this criterion.

Materials and Setting

Two rooms were used: a playroom (4.27 × 3.81 m) and an adjoining observation room separated by a one-way mirror. The playroom was furnished with an armchair, end table, toy chest, small rug, floor pillow, credenza, VCR, and 53.30-cm (21 in., diagonally measured) television set. The toy chest was open shelved and held a variety of age-appropriate toys representative of a standard toy set (e.g., McCune-Nicholich & Fenson, 1984). Pilot research indicated that no one toy in the set was overwhelmingly popular. A variety of current magazines and a newspaper were placed on the end table within reach of the armchair.

During the television session, the VCR played an episode of *Jeopardy!*, a TV game show produced for adult audiences, including the commercials that appeared with the show at the time of airing. The experimenter could control the TV and VCR from the observation room using remote control devices. At the position most frequently occupied by children as they played with toys, the TV screen subtended a visual angle of 17° horizontally. At that position, the audio volume (preset by the experimenter) at head level of a typical 2-year-old child averaged 57dB (C weighting over ten 200-ms samples), about the level of typical human speech.

Two digital video (DV) cameras were used to record the child's behavior. One DV camera, located in the playroom, was hidden in a shoebox covered with black cloth on the floor under the credenza. The other DV camera was mounted on a tripod in the observation room. The two cameras were connected to a DV mini-cassette recorder in the observation room via a switcher. Two separate monitors were also attached to the switcher allowing the experimenter to simultaneously view the perspectives of both cameras. The experimenter could thus choose at any moment which camera captured the best image of the child. An omnidirectional microphone connected to the VCR hung from the ceiling in the playroom.

Procedure

Upon entering the playroom, the children were allowed to become familiar with the toys while the experimenter explained the procedure to the parents. Once written informed consent was obtained, the experimenter left the playroom for the remainder of the session. From the observation room, the experimenter immediately began videotaping the child in the playroom. In the TV first order, the experimenter remotely started the television program as soon as video recording had commenced. When the program ended, after approximately 30 min, the experimenter turned off the television from the observation room and continued to videotape for an additional 30 min without background TV. In the TV second order, the TV remained off for the first 30 min after which the experimenter turned on the TV and VCR with the remote control.

The focus of the present experiment was on background television's influence on children's solitary play. Consequently, parents were asked not to initiate play with their child or to actively encourage their child to play with any particular toy. The experimenter suggested the parents watch TV or read magazines unless their children became fussy or explicitly demanded their attention. The parents were also asked to fill out a questionnaire about their children's media use. The children were free to play with the toys or watch TV for the duration of the hour.

Coding

Videotapes were coded with a computer using the tape-logging utility of the video-editing program, Adobe Premiere. A research assistant watching the videotape, having identified the onset of a specified behavior according to a detailed set of coding instructions, pressed an appropriate key. The computer stored the identity and the video frame number of the event. The research assistant then advanced the tape at any desired speed until the behavior was judged to have ended and pressed an appropriate key. The frame number of the end point was thus stored. In this manner, the temporal onsets and offsets of the behavior stream were recorded.

Coding proceeded in four passes. The first pass recorded the onsets and offsets of children's looks at the television screen. A look began when a child's eyes oriented to the screen and ended when the child looked away (for details, see Anderson & Levin, 1976).

The second pass recorded the onsets and offsets of play episodes. The coding used a procedure developed by Choi and Anderson (1991). An onset was defined as the frame in which a child touches a toy with which the child subsequently plays. An offset is the frame where a child stops actively playing with the toy, even if passive contact is maintained. Simultaneous contact with more than one toy constituted a single play episode if the toys were combined in any way (e.g., banging them together or using several objects in a pretend play sequence such as cooking). In cases where a play episode was interrupted by a look at the TV (when the TV was on), a play episode was coded as ending only if the look lasted longer than 3 s or, if the look was shorter than 3 s, the child did not return to the same toy.

The third pass recorded the onsets and offsets of focused attention during object play for each play episode. Coders were trained to recognize the behavioral characteristics associated with focused attention as described by Ruff and Rothbart (1996). These behaviors included a serious facial expression with furrowed brow, a forward leaning posture toward the object, and minimized extraneous body movements. A similar 3-s rule was employed in that brief looks away from the toy were not coded as ending focused attention provided that the child looked away for less than 3 s and returned to a state of focused attention with the same object.

A fourth pass coded maturity of play using the 12-point scale published by Belsky and Most (1981). This scale was supplemented by two additional levels adapted from an unpublished scale by McCune-Nicholich (1980) to encompass play by somewhat older children (36-month-olds) than had been observed by Belsky and Most. A numerical value was assigned to each level of play. Levels 1 through 12 were exactly as described by Belsky and Most, and Levels 13 and 14 were those added for the purpose of the current study. For instance, "mouthing" or "simple manipulation" (touching a toy), according to the scale by Belsky and Most, received a 1 or 2, respectively. An episode of "functional-relational" play, in which the child combined two objects in a meaningful way (e.g., touching a spoon to a plate), received a 5. A self-directed symbolic play episode received a 7, whereas a play episode involving "substitution" (e.g., the child put a block up to a doll's mouth as though it were a bottle while talking about feeding the doll) received a higher score of 9. Level 13 was used to indicate object agency with descriptive and functional language. Thus, this level designated times in which the child played the role of a doll within a pretend play scheme (e.g., the child attributed dialogue to a plastic Fisher-Price doll). Last, object play in which the child incorporated imaginary objects for which there was no concrete referent received a code of 14.

Play maturity was coded continuously from the beginning to the end of each previously identified play episode. If the maturity of play changed within a play episode, as it often did, coding encompassed that change.

A videotape of a typical child at each age was used for training. A research assistant worked with the training tape until he or she achieved acceptable reliability with an experienced coder (phi correlation greater than .70). At this point, the assistant was asked to code a test tape. The assistant was allowed to code tapes from the study only if he or she attained a high correlation with the supervisor's test tape coding for both the onsets and the offsets of the target behavior.

To assess interobserver reliability (IOR), two independent observers scored 4 participants at each age (12 in all) for each measure. These IOR tapes were distributed throughout the entire coding period to monitor reliability over the course of the study. Agreement was acceptably high for all behaviors. Intraclass correlations were .85 for mean play length, .69 for mean length of focused attention, and .98 for mean level of play maturity weighted by the time spent at each level. Although 12 tapes were double-coded for looks and deemed acceptably high, due to experimenter error, there is only a record of 4 of those double-coded tapes. The phi correlations (a measure of temporal overlap calculated for each pair of observers) for these tapes were all above .90.

Results

The initial analyses for the play measures were 3 (age: 12, 24, and 36 months) × 2 (sex: male and female) × 2 (test order: TV first and TV second) × 2 (condition: TV and no TV) analyses of variance (ANOVAs) with condition as a repeated measure. Dependent measures were percent of session spent in play, mean play episode length, percent of play that was focused, mean length of focused play, and average maturity of play weighted by time spent at each level. An additional between-subjects ANOVA (excluding the condition variable) was run to analyze the total number of looks at the television when it was on.

The ANOVAs revealed order effects for play episodes, focused attention, and the total number of looks at the television. Investigation of these effects indicated that there were systematic curvilinear changes in these measures over time. Therefore, hierarchical linear modeling (HLM) was used to systematically analyze change over time and to isolate the effects of TV and other factors. HLM is a particularly powerful statistical tool for investigating measures that are differentially expressed over time and can simultaneously assess continuous within- and between-subjects factors without overestimating power (for more details on this method, see Raudenbush & Bryk, 2002). Outcome variables for the HLM analyses were measured in 6-min intervals to account for changes over time. For play episodes and focused attention, one set of HLM analyses assessed the effect of TV condition. A second set of analyses assessed the particular influence of looks at the TV screen on play measures during the TV condition. These tests are described in detail within each section below. Because no order effects were found for play maturity, only the simpler results from ANOVA are presented here.

Bivariate correlations between all dependent variables are presented in Table 1. Descriptive statistics for all the dependent measures are presented in Table 2, but the reader should note that these are unadjusted means and as such do not control for variability with time or other factors accounted for in the HLM analyses. A brief description of how to interpret the HLM results can be found in the Appendix. All reported results are significant at alpha level $p < .05$ or better.

Number of Looks at Television

The HLM analysis for number of looks at the TV included linear and quadratic change over time (measured in 6-min intervals) as Level 1 (within-subjects)

Table 1
Bivariate Correlations Between Dependent Variables

	Play percent	Play length	Focused percent	Focused length	Play maturity
Play percent	1.00	.62**	.29*	.47**	.47**
Play length	.62**	1.00	.37**	.63**	.63**
Focused percent	.29*	.37**	1.00	.71**	.38**
Focused length	.47**	.63**	.71**	1.00	.56**
Play maturity	.47**	.63**	.38**	.56**	1.00

$*p < .05$. $**p < .01$.

predictors and child age, sex, and order of conditions as Level 2 (between-subjects) predictors. Looks at the TV were quite short, and visual attention averaged about 5% indicating that television was truly in the background for these children (see Table 2). HLM analyses revealed that the overall number of looks at the TV was greater when the TV condition was presented in the second half hour, producing a significant effect of order, $B = -48.09$ ($SE = 9.65$), $t(47) = -4.98$. The number of looks at the TV generally decreased over time that the TV was on, particularly from the first 6-min interval to the second as indicated by a significant positive quadratic slope, $B = -12.27$ ($SE = 2.59$), $t(48) = -4.73$, $p < .001$, and $B = 0.70$ ($SE = 0.15$), $t(49) = 4.57$, $p < .001$, for linear and quadratic slopes, respectively. The increase in looks at the screen when TV was presented in the second half hour was largely due to more looks in the first 6-min interval that the TV was on, as indicated by a significant effect of order on the time slope, $B = 7.86$ ($SE = 1.69$), $t(48) = 4.66$. Furthermore, looks diminished with age, primarily because 3-year-olds had fewer looks at the TV than the two younger groups (see Table 2; Figure 1), $B = -1.05$ ($SE = 0.28$), $t(47) = -3.73$. Sex was not a significant predictor of the number of looks. The decline in look number from the first 6-min interval to the second suggests that children habituated attention to the TV somewhat during the first 6 min but that there was no further habituation.

Table 2

Means (SEs) for All Dependent Measures as a Function of Age and Condition

	Looking at television			Play episodes		Focused attention		Maturity
	Percent	Number	Length	Percent	Length	Percent	Length	Weighted average
12 months								
TV	5.51 (1.32)	26.55 (2.98)	3.09 (0.31)	65.70 (5.02)	42.57 (7.73)	21.08 (3.84)	12.72 (1.73)	2.65 (0.13)
No TV	—	—	—	71.11 (4.46)	50.65 (6.85)	29.85 (4.97)	17.58 (1.88)	2.60 (0.09)
24 months								
TV	6.15 (1.12)	26.13 (3.06)	3.83 (0.43)	70.82 (3.53)	73.69 (14.14)	36.80 (4.50)	22.29 (2.14)	4.86 (0.35)
No TV	—	—	—	75.37 (3.58)	74.51 (9.79)	31.77 (5.17)	20.60 (3.03)	4.86 (0.41)
36 months								
TV	3.06 (0.68)	16.55 (2.98)	2.93 (0.36)	75.80 (3.35)	72.72 (7.04)	43.59 (5.29)	31.80 (5.00)	5.02 (0.34)
No TV	—	—	—	79.76 (3.13)	106.68 (13.61)	44.81 (5.81)	33.78 (4.32)	5.52 (0.32)
Average								
TV	4.88 (0.64)	23.08 (1.74)	3.27 (0.21)	70.77 (2.37)	62.78 (6.00)	33.44 (2.92)	22.27 (2.19)	4.16 (0.23)
No TV	—	—	—	75.42 (2.19)	77.33 (6.79)	35.55 (3.17)	24.06 (2.10)	4.31 (0.25)

Note. Dependent measures are looking at television (percent of TV condition spent looking at screen, number of looks at TV, and length of looks), play episodes (percent of session spent in play and length of episodes), focused attention (percent of play that was focused and length of focused attention episodes), and play maturity (average maturity of play weighted by time spent at each level). All lengths are measured in seconds. For maturity of play, numbers denote qualitative level of play.

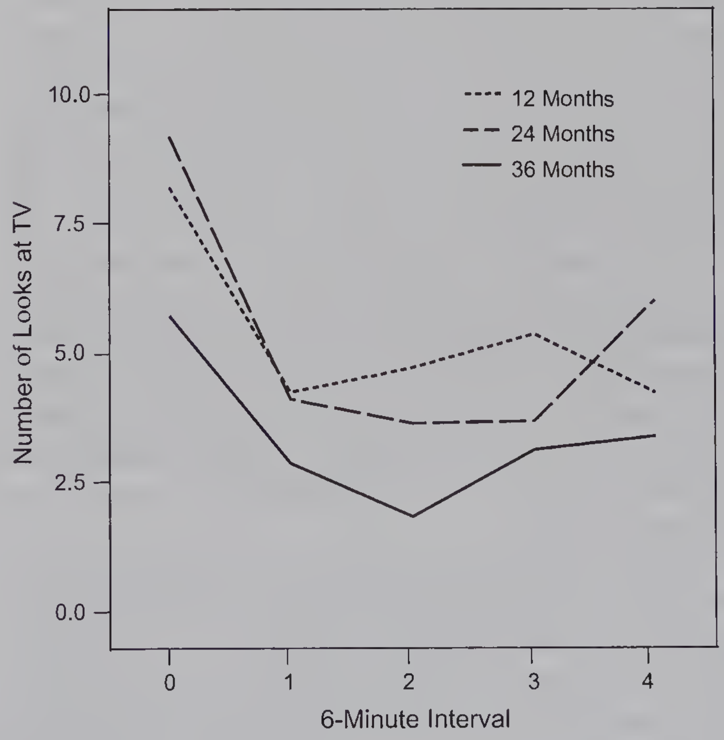


Figure 1. Mean number of looks at TV per 6-min interval as a function of age (values are unadjusted by hierarchical linear modeling coefficients).

Importantly, analyses of the effects of TV on play that excluded data from the first 6-min interval did not affect conclusions resulting from analyses of all the data. That is, the initial novelty of the TV program did not by itself account for any of the results described below. Another important consideration is that because play episodes and episodes of focused attention were coded so that they ended if the child looked at the TV longer than 3 s, it is possible that effects of TV in reducing play or focused attention length are due to these longer looks at the TV (about 30% of all looks). Consequently, we performed analyses considering only looks at the TV that were less than 3 s in length as well as only those longer than 3 s. There were no substantial differences in results from those reported below; that is, looks at the TV that lasted less than 3 s had about the same effects per look on the dependent variables as looks longer than 3 s.

Play and Focused Attention as a Function of Condition

The first set of HLM analyses on play episodes and focused attention considered the dependent measures as a function of condition (TV and no TV). For these analyses, Level 1 (within-subjects) HLM predictors included condition, linear change over time (6-min intervals), quadratic change over time, and interactions between these factors. Level 2 (between-subjects) predictors included age and sex of the child, order of conditions (TV first and TV second), parent education (i.e., highest level achieved), familiarity with the toys in the playroom (i.e., how many of the toys were similar to ones the child had at home), exposure to TV at home (i.e., average of parent-reported hours for each day of the week), and how frequently the child looked at the television in the lab (i.e., total number of looks at the screen as an individual difference measure). All the above factors were assessed as predictors of the outcome measures, but only those that significantly improved the overall model were included in the final models. Specifically, to minimize the chance of Type I error due to multiple tests, only those predictors that significantly contributed to the total amount of variability explained by the model, based on HLM hypothesis testing, are presented here.

Percent of interval spent in play. The percent of each interval spent engaged in toy play decreased over time spent in the room, $B = -0.84$ ($SE = 0.33$), $t(49) = -2.58$. Condition was also a significant Level 1 predictor in the final model for percent of time spent in toy play, $B = -5.08$ ($SE = 1.86$), $t(49) = -2.73$. In this model, condition explained an additional 4.06% of within-subjects (Level 1) variability not accounted for by the model without TV (i.e., after controlling for time). Children played less by about 5 percentage points, or 18 s per 6-min interval, when the TV was on. Because play decreased by about the same amount as the children looked at the TV (i.e., 5% of the session), it seems likely that looking at the TV simply displaced ongoing play.

Length of play episodes. A small decrease in the total amount of play in an interval does not necessarily mean that the length of play episodes decreased. Consequently, play episode length was analyzed separately as a function of condition. The final model for mean play episode length included linear and quadratic change as significant Level 1 predictors, $B = 27.61$ ($SE = 6.95$), $t(46) = 3.97$, and $B = -2.76$ ($SE = 0.76$), $t(46) = -3.64$, for linear and quadratic slopes, respectively. Furthermore, at Level 2, the size of the linear and quadratic slopes was predicted by an interactive combination of age and sex, $B = -23.57$ ($SE = 8.66$), $t(46) = -2.72$, and $B = 3.33$ ($SE = 1.03$), $t(46) = 3.22$, for the Age × Sex interaction on linear and quadratic coefficients, respectively. The overall change over time was such that the average length of play episodes increased until just after the midpoint of the 60-min session and then decreased. The Age × Sex effects were due to 3-year-old girls showing the most dramatic changes over time, whereas younger children and boys showed relatively less change over time. Adjusted for these effects, background television decreased play episode lengths such that play episodes were approximately 30 s shorter on average in the TV condition, $B = -30.16$ ($SE = 7.39$), $t(49) = -4.08$. Condition explained an additional 4.30% of within-subjects variability not accounted for by the model without TV. The only residual variance remaining to be explained by the final model was for the effect of TV suggesting that other factors, such as individual differences not captured by age, sex, toy familiarity, home TV exposure, and parent education, may play a role in the effect of background TV. In sum, play episode lengths changed over time in an inverted-U fashion, particularly for older girls, and became substantially shorter in the presence of background television when controlling for these general changes over time.

Percent of play that was focused. Because background TV reduced the total amount of play, we analyzed the percent of play that was accompanied by focused attention. The final model for percent of focused play included only linear change as a predictor at Level 1. Significant Level 2 predictors in the final model included the effect of age on the intercept and an Age × Sex interaction on the slope for time. The amount of play that was focused increased by about 11% per year of age, B = 10.98 (SE = 3.99), t(48) = 2.75. In addition, the percent of play that was focused decreased linearly over time, B = −0.82 (SE = 0.38), t(46) = −2.16. This was particularly true of males and younger females, B = −1.59 (SE = 0.63), t(46) = −2.52. The effect of TV condition and the quadratic slope was not significant for the percent of play that was focused, although significant variability remained to be explained suggesting that changes occurred in different directions for individual children. No Level 2 variables in the current study were able to sufficiently predict these changes.

Length of focused episodes. Even though the percent of focused play changed little in the presence of TV, the lengths of episodes of focused attention were affected. The final model for all data for mean length of focused episodes included linear and quadratic change over time and condition at Level 1. Similar to the results for mean play episode length, in this model, the quadratic slope was negative such that initially length of focused attention episodes increased until just after the midpoint and then decreased, B = −0.40 (SE = 0.19), t(48) = −2.13. Furthermore, age was a significant Level 2 predictor of the quadratic slope insofar as the changes over time were more dramatic for older children, B = −0.65 (SE = 0.17), t(48) = −3.87. Adjusted for these effects, television decreased the mean length of focused attention episodes by approximately 5 s (almost 25%) during the first interval, B = −5.06 (SE = 2.17), t(49) = −2.33. Condition accounted for an additional 5.73% of within-subjects variability. There was significant variability remaining to be explained for change over time and the effect of TV. Again, these are likely due to unexplained individual differences.

Play and Focused Attention in the Presence of TV as a Function of Looks at the Screen

The second set of HLM analyses was conducted using only the data from the TV sessions for each subject. Rather than assessing play behavior as a function of condition, in this second set of analyses, the number of looks at the TV screen was included as a Level 1 predictor to determine the extent to which this measure of distraction by the TV predicted variation in outcome measures when the TV was on. In other words, for each participant, the outcome measure within a 6-min interval (e.g., percent of that interval spent in play) was assessed as a function of the number of looks at the TV that occurred during that same 6-min interval. Other within-subjects predictors were linear and quadratic change over time and interactions between time and looks at TV. Between-subjects factors at Level 2 included order of conditions, parent education, toy familiarity, and home TV exposure.

Percent of interval spent in play. When considering data from the TV condition alone, the amount of time spent in play decreased significantly over time, B = −2.21 (SE = 0.97), t(49) = −2.28. In addition, the percent of the interval spent in play significantly decreased by an average 2.3% with each additional look at the TV screen in an interval (see Figure 2), B = −2.31 (SE = 0.38), t(49) = −6.15. Children spent more time engaged in toy play during intervals in which they had relatively few looks at the TV screen as compared to intervals during which they had more looks at the screen. The number of looks at the screen accounted for 14.27% of within-subjects variability remaining to be explained after controlling for change over time. Similar to findings reported above, the negative effect of look number on play behavior held even when excluding data from the first 6 min of the session during which looks at the TV were particularly high as well as when considering only those looks that were less than 3 s in length.

Length of play episodes. The number of looks per 6-min interval was also a significant negative predictor of the length of play episodes in the TV condition (see Figure 3), B = −3.32 (SE = 0.94), t(49) = −3.54. For each additional look at the TV in an interval, the mean length of play episodes in that interval dropped by an

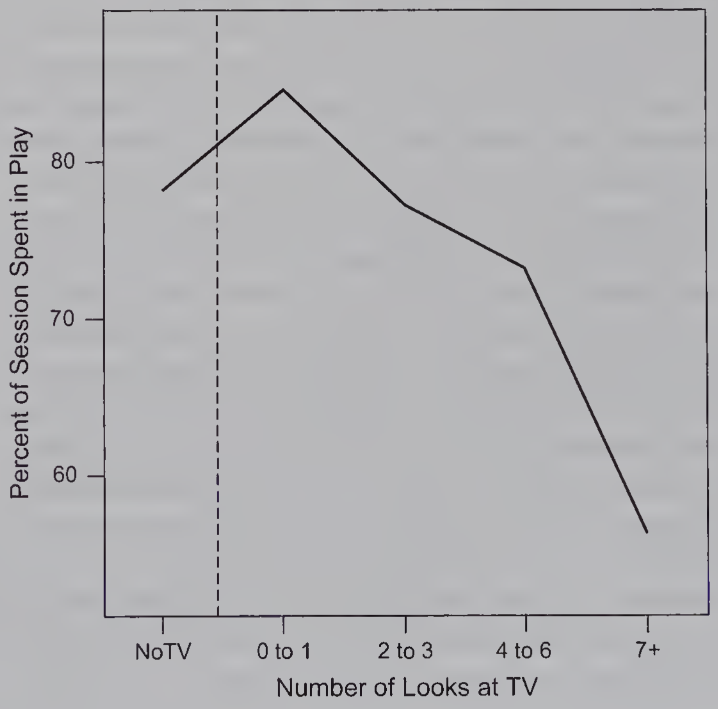


Figure 2. Mean percent of 6-min intervals spent in play as a function of the number of looks at the TV.
Note. Values are unadjusted by hierarchical linear modeling coefficients. Average percent of session spent in play for the no TV condition is provided for reference.

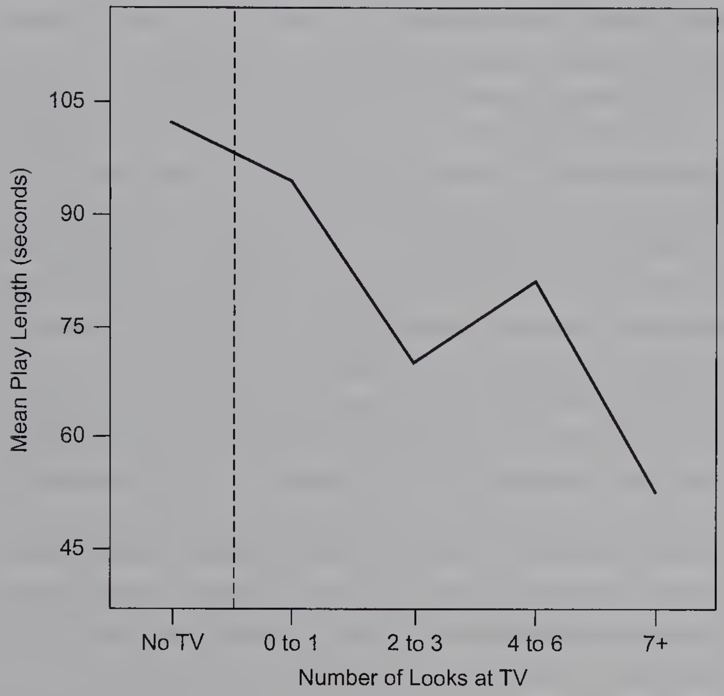


Figure 3. Mean length of play episodes per 6-min interval during the TV condition as a function of the number of looks at TV during that interval (values are unadjusted by hierarchical linear modeling coefficients).

Note. Average episode length during the no TV condition is provided for reference.

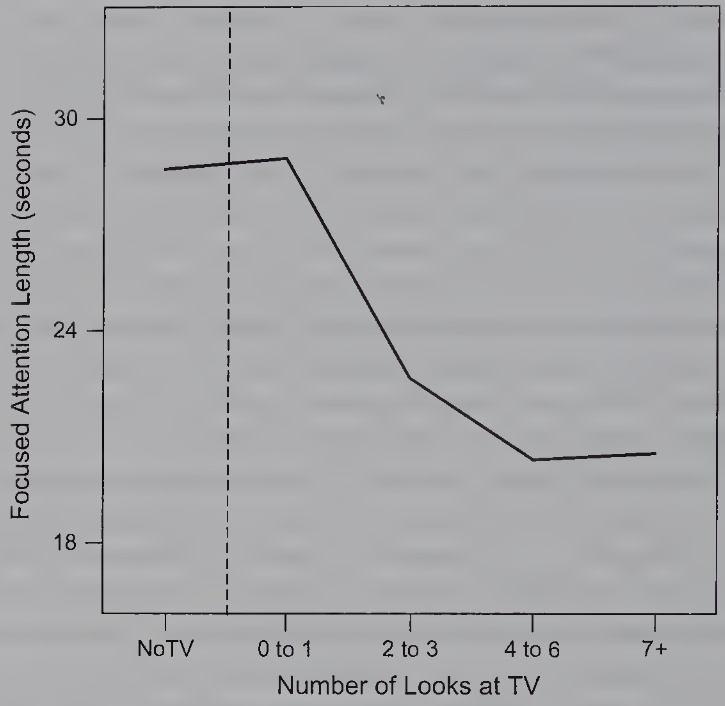


Figure 4. Mean length of focused attention episodes per 6-min interval during the TV condition as a function of the number of looks at TV during that interval.

Note. Average episode length during the no TV condition is provided for reference. Means are unadjusted by hierarchical linear modeling coefficients.

average of 3.3 s. The number of looks at the screen per interval accounted for an additional 4.57% of within-subjects variability.

Percent of play that was focused. As with the HLM analysis on the entire session, the percent of play in the TV condition that was focused increased with age, $B = 10.88$ ($SE = 3.07$), $t(48) = 3.55$. Although the percent of play that was focused in the TV condition did decrease somewhat as a function of the number of looks at the screen, this effect was not significant.

Length of focused play episodes. The quadratic slope was significant reflecting the same inverted-U changes over time that were found in the first analysis on length of focused episodes, $B = -13.41$ ($SE = 3.56$), $t(48) = -3.77$, and $B = 3.20$ ($SE = 0.79$), $t(48) = 4.06$, for linear and quadratic slopes, respectively. Furthermore, the number of looks at the TV was a significant predictor of the length of focused attention as shown in Figure 4, $B = -1.14$ ($SE = 0.45$), $t(48) = -2.53$. The length of focused attention episodes decreased by an average 1.14 s per look at the TV. The number of looks at the screen accounted for 19.91% of within-subjects variability remaining to be explained.

Maturity of Toy Play

Weighted average play maturity. Because exploratory analyses of maturity of toy play indicated that there were no order effects indicative of systematic changes over time, the results presented are from mixed ANOVAs. For average play maturity (weighted by the time spent at each level of maturity), the analysis was a 3 (age) × 2 (sex) × 2 (order of conditions) × 2 (condition) ANOVA with condition as a repeated measure. The only significant main effect for average maturity level was age, $F(2, 38) = 42.87$. Post hoc analyses revealed that the 12-month-olds (mean level = 2.62) differed significantly from the two older age groups but that the 24- and 36-month-olds were not significantly different from each other (mean levels = 4.86 and 5.28, respectively). There was also a significant interaction between age and order of conditions, $F(2, 38) = 4.90$. Average maturity of play by 36-month-olds was greater when the TV condition was presented first than when it was second. There were no significant main effects or interactions with respect to the presence or absence of background television.

Time spent at each level of play. Exploration of the play maturity data indicated that the children spent the vast majority of their play at only 4 of the 14 levels: Level 2 (simple manipulation), Level 3 (functional), Level 5 (functional-relational), and Level 10 (sequence pretend). Level 3 refers to appropriate play with an object according to its standard intended use. An example is turning the crank on a jack-in-the-box.

Level 5 refers to play that puts two or more objects together in an appropriate manner (e.g., putting lids on pans). Level 10 refers to play that involves pretend acts in a related series (e.g., picking up doll, getting bottle, feeding doll with bottle, putting doll in bed). Together, these four categories accounted for 84, 82, and 83% of total play for 12-, 24-, and 36-month-olds, respectively. An ANOVA was conducted to determine whether TV had a differential effect on any of these four levels of play maturity. Time spent at each of these four levels was subjected to a 3 (age) × 2 (sex) × 2 (order) × 2 (condition) × 4 (level of play: 2, 3, 5, and 10) ANOVA with condition and level as repeated measures.

The omnibus ANOVA revealed a significant Condition × Level interaction that was explored by ANOVA on each level independently. Because of the large number of subordinate tests run, a more stringent significance level of .01 was adopted. There was a significant main effect of age for Levels 2, 5, and 10, $F(2, 38) = 40.89$, $F(2, 38) = 16.79$, and $F(2, 38) = 14.44$, respectively. Post hoc analyses (*t* tests with Bonferroni correction) revealed that 12-month-olds engaged in significantly more play at Level 2 (51% of play vs. 24% and 18%) and significantly less play at Levels 5 (8% vs. 22 and 31%) and 10 (0.5% vs. 14 and 16%) than 24- and 36-month-olds. The two older age groups did not differ significantly at any level.

The only level of play maturity demonstrating a significant effect of TV was Level 3 (functional play). For functional play, there was a significant Condition × Order interaction such that there was more play at this level in the second session compared to the first, particularly when TV was presented second, $F(1, 38) = 16.80$. In the first half hour, Level 3 accounted for 17% of play for children with TV and 16% for children without TV. In the second half hour, children with TV spent 31% of play at Level 3, whereas children without TV spent 24% of play at Level 3. A compensatory decrease in Level 5 play in the presence of TV approached significance, $F(1, 38) = 5.19$.

Discussion

We found that very young children's toy play was disrupted by background television. Specifically, compared to no television, there was less play overall, shorter play episodes, and shorter bouts of focused attention in the presence of background television. Given these findings, it was surprising that there was little overall reduction in focused attention and that maturity of play was affected in a limited way. Although there was some reduction in functional-relational play in the presence of television with a compensatory increase in functional play, there was no reduction in complex, high-level play as predicted. Overall, the disruptive effects are real but small.

We hypothesized that background television would disrupt play by repeatedly eliciting an OR toward the television screen and away from ongoing play activities. The hypothesis was supported insofar as children frequently oriented toward the TV screen for short periods of time, although there was a reduction of such looks after the first 6 min of the program. The more frequently the child looked at the TV screen, the greater was the disruption of play.

The overall reduction of play in the presence of background television is likely due to displacement by looking at the screen. The cumulative time spent playing was reduced by about 5% in the half hour that the television was on, just about the same as the 5% of time spent looking at the television. Although an important finding on its own, this reduction in overall time spent playing is not sufficient to account for the reduced length of play episodes. The overall reduction in play also does not account for the reduction in lengths of focused attention or the slight reduction in play maturity. It is likely that in very young children, brief looks at background television disrupt ongoing play schemes. When the child stops looking at the TV, he or she may have forgotten the ongoing play scheme or otherwise find it difficult to reinstantiate the scheme in which case a new scheme is initiated with another toy. In fact, in the presence of television, children were more likely to move from toy to toy, coming into contact with an average of 65% of the toys as compared to 55% when the television was not on.

Background television has been shown to reduce adults' performance on a variety of complex cognitive tasks. This has been interpreted as the television competing with the primary task for cognitive resources (e.g., Armstrong & Greenberg, 1990). Assuming that toy play is cognitively demanding in very young children, interference with play could be expected as well. Background television may also be implicated in a more general kind of interference associated with noisy environments. Children exposed to excessive noise (e.g., from airports or motor traffic) do less well on tests and on tasks requiring concentration (Hygge et al., 2002). With respect to very young children, the level of ambient noise in the home is negatively related to cognitive development (Gottfried, 1984; Wachs, 1985, 1986; Wachs & Gandour, 1983). In a related vein, Landry, Smith, Swank, and Miller-Loncar (2000) found that the extent to which parent interactions disrupt children's ongoing play negatively predicts children's subsequent cognitive and language skills. These authors proposed that interrupting children's play requires them to repeatedly refocus attention and may tax children's limited cognitive resources. Background television may have an analogous effect.

This research raises the question of the influence of background television in children's natural environments. Assuming that the effects found in this experiment occur at home, it is as yet unknown whether chronic reductions in play episode lengths and focused attention lengths produced by background television are parts of a causal chain leading to negative long-term consequences. Nevertheless, short play episode lengths are a marker for poor developmental outcome (e.g., Faden & Graubard, 2000; Handen et al., 1998; Malone, 1997) as is reduced focused attention during play (Lawson & Ruff, 2004). Moreover, children's sustained attention has been shown to mediate effects of family environment (composite of physical environment and caregiver responsiveness) on school readiness (National Institute of Child Health and Human Development Early Child Care Research Network, 2003).

Although the present research shows that background television plays a proximate causal role in reducing play episode length and focused attention, it remains for future research to determine whether this suite of effects is by itself causal in poorer developmental outcomes. Prior research on background noise in the home, of which background television is a considerable component, has identified a negative relationship between such noise and mothers' verbal responsiveness to the child (Corapci & Wachs, 2002). This suggests that background television may exert an influence beyond its direct effects on very young children's solitary toy play as observed in the present work. Indeed, in a subsequent experiment following a design parallel to the present study, we found that the quality and quantity of parent–child interactions are also affected as parents watch or become distracted by television (Kirkorian, Murphy, Pempek, Anderson, & Schmidt, 2005). It is also possible that chronic exposure to television that is constantly left on in the home is correlated with other factors, such as a generally chaotic home environment (Wachs, 1986). Taken together, however, reduced parent–child interactions, combined with reduced focused attention during play and reduced play episode length, lead us to hypothesize that background television, as a chronic influence, is by itself an environmental risk factor in children's development.

It is possible, of course, that background television does not induce effects that are necessarily negative but instead induces a characteristic style of attention deployment, a style that may be appropriate in some situations but not in others. Correa-Chavez, Rogoff, and Arauz (2005), for example, argue that the attentional style of traditional village-reared Latin American children differs from that of European American children with parents who have received extensive schooling. These distinctions (attending sequentially vs. simultaneously to multiple information sources) are produced by substantially different childhood experiences and may prove optimal in different situations (e.g., learning in school vs. observationally learning from village activities). From this point of view, background TV may induce an attentional style that is not necessarily negative in all situations. For example, it is possible that it can provide some basis for later multitasking skills. Contemporary American youth are characterized as simultaneously talking on cell phones and playing computer games even as they monitor TV (Foehr, 2006). It has not yet been shown whether such media multitasking is associated with generally positive or negative outcomes.

Our basic conception of background television is that it functions as a continuously varying distractor for very young children. Whereas it is thought that distractibility decreases with age, and the evidence supports this in school-aged children (e.g., Strutt, Anderson, & Well, 1975), Ruff and Rothbart (1996) note that there is little evidence for a developmental trend in distractibility for children aged 5 years and younger. In the present study, if looking away from toy play to the television is taken as a measure of distraction, the 3-year-olds looked fewer times than younger children. Consistent with Ruff and Rothbart's model of attention development, play length and focused attention also increased with age. However, we did not find an age difference in the disruptive effects of background television, a continuously changing audiovisual distractor. Rather, all ages had reduced play episode and focused attention lengths.

It is possible that these effects are uniquely produced by television as a distractor. Ruff and Capozzoli (2003) examined the effects of an intermittent audiovisual distractor on the toy play of 10-, 26-, and 42-month-olds. Like the present study, they found a decrease in distractibility over time; their experimental session was 12 min in length. However, they found no difference in distractibility as a function of age as evidenced by looks at the distractor. Ruff and Capozzoli (2003), moreover, found that the intermittent distractor did not reduce either length or amount of focused attention during play. It could be that the difference between studies is due to the continuously changing nature of background television, but the differences might also be due to several factors including the longer observation period, a greater number of toys in the present experiment, general age differences, or other procedural differences.

In conclusion, even though the effects of background television on play behavior found in this study are small, they may have a cumulative impact through large amounts of exposure at home. These may include poorer cognitive and language development and attention deficit symptoms as found in associations from a small number of studies that focus on very young children (Carew, 1980; Christakis et al., 2004; Nelson, 1973; Wachs, 1985, 1986; Zimmerman & Christakis, 2005). Researchers to date, lacking direct evidence, have tended to attribute these associations to negative effects of television programs made for young children and to which they are highly attentive (e.g., Christakis et al., 2004). The present research, in contrast, is suggestive that programs made for adults and watched by parents may in fact provide a risk factor for development and could conceivably account for the negative associations with cognitive and language development. This study demonstrates that it is imperative for future research on the influence of media in the home to distinguish between the possibly greatly differing effects of foreground and background television exposure.

References

American Academy of Pediatrics, Committee on Public Education. (1999). Media education. *Pediatrics*, *104*, 341–342.

Anderson, D. R., & Evans, M. K. (2001). Peril and potential of media for infants and toddlers. *Zero to Three: National Center for Infants, Toddlers, and Families*, *22*, 10–16.

Anderson, D. R., & Levin, S. R. (1976). Young children's attention to *Sesame Street*. *Child Development*, *47*, 806–811.

Anderson, D. R., Lorch, E. P., Collins, P. A., Field, D. E., & Nathan, J. G. (1986). Television viewing at home: Age trends in visual attention and time with TV. *Child Development*, *57*, 1024–1033.

Anderson, D. R., Lorch, E. P., Field, D. E., & Sanders, J. (1981). The effect of television program comprehensibility on preschool children's visual attention to television. *Child Development*, *52*, 151–157.

Armstrong, G. B., & Greenberg, B. S. (1990). Background television as an inhibitor of cognitive processing. *Human Communication Research*, *16*, 355–386.

Belsky, J., & Most, R. K. (1981). From exploration to play: A cross-sectional study of infant free play behavior. *Developmental Psychology*, *17*, 630–639.

Bretherton, I. (1984). *Symbolic play: The development of social understanding*. Orlando, FL: Academic Press.

Carew, J. V. (1980). Experience and the development of intelligence in young children at home and in day care. *Monographs of the Society for Research in Child Development*, *45*(Serial No. 187).

Certain, L. K., & Kahn, R. S. (2002). Prevalence, correlates, and trajectory of television viewing among infants and toddlers. *Pediatrics*, *109*, 634–642.

Choi, H. P., & Anderson, D. R. (1991). A temporal analysis of toy play and distractibility in young children. *Journal of Experimental Child Psychology*, *52*, 41–69.

Christakis, D. A., Zimmerman, F. J., DiGiuseppe, D. L., & McCarty, C. A. (2004). Early television exposure and subsequent attentional problems in children. *Pediatrics*, *113*, 708–713.

Clarke-Stewart, K. A. (1973). Interactions between mothers and their young children: Characteristics and consequences. *Monographs of the Society for Research in Child Development*, *38*(6–7, Serial No. 153).

Corapci, F., & Wachs, T. D. (2002). Does parental mood or efficacy mediate the influence of environmental chaos upon parenting behavior? *Merrill-Palmer Quarterly*, *48*, 182–201.

Correa-Chavez, M., Rogoff, B., & Arauz, R. M. (2005). Cultural patterns in attending to two events at once. *Child Development*, *76*, 664–678.

Derryberry, D., & Tucker, D. M. (1990). The adaptive base of the neural hierarchy: Elementary motivational controls on network function. In R. Dienstbier (Ed.), *Nebraska Symposium on Motivation: Vol. 38. Perspectives on motivation* (pp. 289–342). Lincoln: University of Nebraska Press.

DiLalla, L. F., & Watson, M. (1988). Differentiation of fantasy and reality: Preschooler's reactions to interruptions in their play. *Developmental Psychology*, *24*, 286–291.

Faden, V. B., & Graubard, B. I. (2000). Maternal substance use during pregnancy and developmental outcome at age three. *Journal of Substance Abuse*, *12*, 329–340.

Foehr, U. G. (2006). *Media multitasking among American youth: Prevalence, predictors, and pairings*. Menlo Park, CA: The Henry J. Kaiser Family Foundation.

Gertner, J. (2005, April 10). Our ratings, ourselves. *The New York Times Magazine*, pp. 34–41.

Gottfried, A. W. (Ed.). (1984). *Home environment and early cognitive development: Longitudinal research*. Orlando, FL: Academic Press.

Handen, B. L., McAuliffe, S., Janosky, J., Feldman, H., & Breaux, A. M. (1998). A playroom observation procedure to assess children with mental retardation and ADHD. *Journal of Abnormal Child Psychology*, *26*, 269–277.

Hygge, S., Evans, G. W., & Bullinger, M. (2002). A prospective study of some effects of aircraft noise on cognitive performance of schoolchildren. *Psychological Science*, *13*, 469–474.

Kirkorian, H. L., Murphy, L. A., Pempek, T. A., Anderson, D. R., & Schmidt, M. E. (2005, April). *The impact of background television on parent-child interaction*. Poster session presented at the biannual meeting of the Society for Research in Child Development, Atlanta, GA.

Landry, S. H., Smith, K. E., Swank, P. R., & Miller-Loncar, C. L. (2000). Early maternal and child influences on children's later independent cognitive and social functioning. *Child Development, 71,* 358–375.

Lawson, K. R., & Ruff, H. A. (2004). Early focused attention predicts outcome for children born prematurely. *Journal of Developmental and Behavioral Pediatrics, 25,* 399–406.

Lemish, D. R. (1987). Viewers in diapers: The early development of television viewing. In T. R. Lindlof (Ed.), *Natural audiences: Qualitative research of media uses and effects* (pp. 35–57). Norwood, NJ: Ablex.

Linebarger, D. L., & Walker, D. (2005). Infants' and toddlers' television viewing and language outcomes. *American Behavioral Scientist, 48,* 624–645.

Malone, D. M. (1997). Preschoolers' categorical and sequential toy play: Changes over time. *Journal of Early Intervention, 21,* 45–61.

McCall, R. B. (1974). Exploratory manipulation and play in the human infant. *Monographs of the Society for Research in Child Development, 39*(2, Serial No. 155).

McCune-Nicholich, L. (1980). *A manual for analyzing free play.* Unpublished manuscript.

McCune-Nicholich, L., & Fenson, L. (1984). Methodological advances in the study of early pretend play. In T. D. Hawkey & A. D. Pellegrini (Eds.), *Child's play* (pp. 81–104). Hillsdale, NJ: Erlbaum.

National Institute of Child Health and Human Development: Early Child Care Research Network. (2003). Do children's attention processes mediate the link between family predictors and school readiness? *Developmental Psychology, 39,* 581–593.

Nelson, K. (1973). Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development, 38*(1–2, Serial No. 149).

Obel, C., Henrickson, T. B., Dalsgaard, S., Linnet, K. M., Skajaa, E., Thomsen, P. H., et al. (2004). Does children's watching of television cause attention problems? Retesting the hypothesis in a Danish cohort. *Pediatrics, 114,* 1372–1373.

Piaget, J. (1962). *Play, dreams and imitation in childhood.* New York: Norton.

Pierroutsakos, S. L., Hanna, M. M., Self, J. A., Lewis, E. N., & Brewer, C. J. (2004, May). *Baby Einsteins everywhere: The amount and nature of television and video viewing of infants birth to 2 years.* Presented at the biennial International Conference for Infant Studies, Chicago.

Power, T. G. (2000). *Play and exploration in children and animals.* Mahwah, NJ: Erlbaum.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Reeves, B., Thorson, E., & Schleuder, J. (1986). Attention to television: Psychological theories and chronometric measures. In J. Bryant & D. Zillman (Eds.), *Perspectives on media effects* (pp. 251–280). Hillsdale, NJ: Lawrence Erlbaum.

Richards, J. E., & Casey, B. J. (1992). Development of sustained visual attention in the human infant. In B. A. Campbell, H. Hayne, & R. Richardson (Eds.), *Attention and information processing in infants and adults* (pp. 30–60). Hillsdale, NJ: Erlbaum.

Richards, J. E., & Cronise, K. (2000). Extended visual fixation in the early preschool years: Look duration, heart rate changes, and attentional inertia. *Child Development, 71,* 602–620.

Richards, J. E., & Gibson, T. L. (1997). Extended visual fixation in young infants: Look distributions, heart rate changes, and attention. *Child Development, 68,* 1041–1056.

Rideout, V., & Hamel, E. (2006). *The media family: Electronic media in the lives of infants, toddlers, preschoolers and their parents.* Menlo Park, CA: The Henry J. Kaiser Family Foundation.

Rideout, V., Vandewater, E., & Wartella, E. (2003). *Zero to six: Electronic media in the lives of infants, toddlers, and preschoolers.* Menlo Park, CA: The Henry J. Kaiser Family Foundation.

Ruff, H. A., & Capozzoli, M. C. (2003). Development of attention and distractibility in the first 4 years of life. *Developmental Psychology, 39,* 877–890.

Ruff, H. A., Capozzoli, M., & Saltarelli, L. M. (1996). Focused visual attention and distractibility in 10-month-old infants. *Infant Behavior & Development, 19,* 281–293.

Ruff, H. A., & Lawson, K. R. (1990). Development of sustained, focused attention in young children during free play. *Developmental Psychology, 26,* 85–93.

Ruff, H. A., & Rothbart, M. K. (1996). *Attention in early development: Themes and variations.* New York: Oxford University Press.

Ruff, H. A., Saltarelli, L. M., Capozzoli, M., & Dubiner, K. (1992). The differentiation of activity in infants' exploration of objects. *Developmental Psychology, 28,* 851–861.

Schmitt, K. L., Anderson, D. R., & Collins, P. A. (1999). Form and content: Looking at visual features of television. *Developmental Psychology, 35,* 1156–1167.

Schmitt, K. L., Woolf, K. D., & Anderson, D. R. (2003). Viewing the viewers: Viewing behaviors by children and adults during television programs and commercials. *Journal of Communication, 53,* 265–281.

Stevens, T., & Mulsow, M. (2006). There is no meaningful relationship between television viewing and symptoms of attention deficit/hyperactivity disorder. *Pediatrics, 117,* 665–672.

Strutt, G. F., Anderson, D. R., & Well, A. D. (1975). A developmental study of the effects of irrelevant information on speeded classification. *Journal of Experimental Child Psychology, 20,* 127–135.

Wachs, T. D. (1985). Home stimulation and cognitive development. In C. C. Brown & A. W. Gottfried (Eds.), *Play interactions: The role of toys and parental involvement in children's development.* (pp. 142–152). Skillman, NJ: Johnson & Johnson Baby Products.

Wachs, T. D. (1986). Ambient background noise and early development. *Children's Environments Quarterly, 3*(1), 23–33.

Wachs, T. D., & Gandour, M. (1983). Temperament, environment and 6 months cognitive intellectual development. *International Journal of Behavioral Development*, *6*, 135–152.

Wachs, T. D., & Gruen, G. E. (1982). *Early experience and human development*. New York: Plenum Press.

Woodard, E. (2000). *Media in the home 2000: The fifth annual survey of parents and children. Survey no. 7*. Philadelphia: The Annenberg Public Policy Center of the University of Pennsylvania.

Zimmerman, F. J., & Christakis, D. A. (2005). Children's television viewing and cognitive outcomes. *Archives of Pediatrics and Adolescent Medicine*, *159*, 619–625.

Zimmerman, F. J., Christakis, D. A., & Meltzoff, A. N. (2007). Television and DVD/video viewing in children younger than two years. *Archives of Pediatrics and Adolescent Medicine*, *161*, 473–479.

Appendix

	Coefficient (*SE*)	Significance test
Constant		
Intercept	57.32 (10.76)	*t*(47) = 5.36, *p* < .001
Age	−1.05 (0.28)	*t*(47) = −3.73, *p* = .001
Order	−48.09 (9.65)	*t*(47) = −4.98, *p* < .001
Time		
Intercept	−12.27 (2.59)	*t*(48) = −4.73, *p* < .001
Order	7.86 (1.69)	*t*(48) = 4.66, *p* < .001
Time squared		
Intercept	0.70 (0.15)	*t*(49) = 4.57, *p* < .001

The table in this Appendix presents the HLM final model for the number of looks at the screen in the TV condition. Order is a dichotomous variable such that a value of 1 represents the TV first order, whereas a value of 0 represents the TV second order. Centering, or transforming variables such that a value of 0 represents a point (actual or estimated) in a data set, is a common practice in HLM analyses. This allows the constant (i.e., the value of the outcome measure when all other variables are set to 0) to be a meaningful value. In this case, age is represented in years and centered at its mean such that a value of 0 represents age 2 years, 1 year of age is represented by a value of −1, and 3 years of age is represented by a value of 1. Time is centered at the first interval of the TV condition such that a value of 0 represents the average for the first 6 min that the TV was on, a value of 1 represents the second 6-min interval, and so on.

Coefficients for "intercept" are the intercepts of the Level 2 equations; coefficients for all other factors are the slopes associated with those predictors. Level 1 (within-subjects) variables can be interpreted in much the same way as in regression analyses. Level 2 (between-subjects) predictors can be interpreted as regression coefficients for the Level 1 slopes. The constant represents the value of the outcome measures when all predictors are set to 0 (the intercept in a standard regression equation). For example, based on the above model, the average number of looks at the screen in the first interval of the TV condition (Time 0) for 2-year-olds (age 0) when TV was presented in the second half hour (Order 0) was about 57. When TV was presented in the first half hour (i.e., when order has a value of 1), the constant is reduced by about 48, resulting in an average of only 9 looks in the first interval for 2-year-olds. The negative slope for age on the constant indicates that the average number of looks decreased with age (approximately one look per 6-min interval per year of age).

There was also a significant negative linear component, indicating that the number of looks generally decreased over the course of the TV condition. Order is a positive predictor of that slope. When TV is presented in the second half hour (i.e., when order is 0), the linear slope is about −12.3; when TV is presented first (i.e., when order has a value of 1), the linear slope increases by a factor of about 7.9, making it only −4.4 for this order. Thus, the number of looks generally decreased with time that the TV was on but less so when TV was presented first than when it was presented second. Last, the positive slope for time squared suggests a U-shaped function with respect to time; this is manifest in a large decrease in looks from the first interval to the second and relatively little change over time thereafter.

Child Development, July/August 2008, Volume 79, Number 4, Pages 1152–1167

German Children's Comprehension of Word Order and Case Marking in Causative Sentences

Miriam Dittmar
Max Planck Institute for Evolutionary Anthropology and University of Zurich

Kirsten Abbot-Smith
University of Plymouth

Elena Lieven and Michael Tomasello
Max Planck Institute for Evolutionary Anthropology

Two comprehension experiments were conducted to investigate whether German children are able to use the grammatical cues of word order and word endings (case markers) to identify agents and patients in a causative sentence and whether they weigh these two cues differently across development. Two-year-olds correctly understood only sentences with both cues supporting each other—the prototypical form. Five-year-olds were able to use word order by itself but not case markers. Only 7-year-olds behaved like adults by relying on case markers over word order when the two cues conflicted. These findings suggest that prototypical instances of linguistic constructions with redundant grammatical marking play a special role in early acquisition, and only later do children isolate and weigh individual grammatical cues appropriately.

One of the important tasks of early childhood is mastering a conventional language. Languages differ not only in their words but also in the grammatical constructions they employ for assembling words into meaningful utterances. Grammatical constructions are composed of multiple words, or word categories, structured into patterns in particular ways by such things as word order and grammatical markers (e.g., a different ending on a word when it is the subject rather than the direct object in a sentence—so-called case marking). In English, the sentence "The dax mibbed the gazzer a toma" (the ditransitive construction) implies a transfer of some kind, even though all the contentful words are meaningless (Goldberg, 1995).

One construction of particular importance in early development is the basic transitive construction, prototypically used to indicate an agent causally acting on an object, as in simply "The dax mibbed the gazzer." The importance of this construction stems from the fact that it is one of the ontogenetically earliest in which it is critical to distinguish the different roles of the participants in some event. Thus, "The toma mibbed" creates no problems for deciding who was doing the action because there is only 1 participant. But if we hear "The toma the gazzer mibbed," we must decide who is mibbing whom, and to do this we need to understand the grammatical conventions of the particular language being learned. Interestingly, in most languages, there are multiple, redundant cues for helping the listener do this in many utterances—although in other utterances, there can be just a single cue. For example, in the English sentence "He mibs pencils," we identify the agent of the action as *he* based on the facts that: (a) it is said before rather than after the action word or verb (word order), (b) it is the subject pronoun *he* (and not the object pronoun *him*; case marking), (c) it agrees in number with the verb (we say "He mibs" but "Pencils mib," without an -s; subject–verb agreement), and (d) it is a statistical fact that animate beings, such as male persons, are more likely to act on inanimate things, such as pencils, than the other way around (animacy). A child acquiring the English transitive construction, therefore, could on a particular occasion be using any or all these cues to determine who is mibbing whom in the utterance.

Many thanks to Elena Rossi, Lena Verfürth, Eva Leermann, Jan Jurkat, and Julia Ulber for assistance in testing and for organizing appointments; to Silke Brandt, Susanne Grassmann, and Catherine van Maanen for coding reliabilities; to Petra Jahn for technical assistance; and to Daniel Stahl for help with counterbalancing and statistics. A special thank you to the parents and children who participated in the study.

Correspondence concerning this article should be addressed to Miriam Dittmar, University of Zurich, Institute of Psychology, Binzmuehlestrasse 14/21, CH-8050 Zurich, Switzerland. Electronic mail may be sent to m.dittmar@psychologie.uzh.ch.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7904-0022

The competition model of Bates and MacWhinney (1987, 1989) represents an attempt to assess how children acquire the different cues of different languages—especially in the transitive construction—and how they weigh these cues relative to one another when they conflict (see also Slobin & Bever, 1982). In a comprehension task in which they were asked to identify who was doing what to whom, young children heard a sentence such as "Him kissed she." Case-marked pronouns indicate that the female kissed the male (she > him), but word order indicates that the male kissed the female (him > she). The finding is that from early in the preschool period English-speaking children privilege word order over all other tested cues (e.g., animacy and subject–verb agreement) in interpreting transitive sentences. Other researchers have tested English-speaking children's comprehension of word order when it is the only cue available (therefore not conflicting with any other cues) and found that even young 2-year-olds already distinguish between such things as "X is tickling Y" and "Y is tickling X" (Hirsh-Pasek & Golinkoff, 1996, with familiar verbs; Gertner, Fisher, & Eisengart, 2006, with novel verbs).

In many other languages, the grammatical cues in transitive sentences are much more evenly weighted than in word-order dominant English. For example, various languages in which all nouns are case marked for their role in the sentence (not just pronouns, as in English), have a much more flexible word order—because if a word is locally marked with a case marker indicating its role in the sentence, then word order may be used for pragmatic functions such as emphasis and perspective (as English does awkwardly in such sentences as "**Him** I like"). Therefore, if German adults are presented with a sentence parallel to the English sentence above ("Him kissed she"), they interpret it in the opposite way to English adults, that is, they insist that the subject-marked pronoun *she* indicates the one doing the kissing even though it comes after the verb (whereas it most often comes before the verb) (MacWhinney, Bates, & Kliegl, 1984).

Importantly, in the competition model, there are methods for quantifying the strength of various cues in a particular language, for example, in the transitive construction. To do this one looks at the general dimensions of the frequency of a cue (cue availability), the consistency of a cue in indicating a function (cue reliability), and the complexity of a cue (cue cost). In English, case-marked pronouns (e.g., *I–me, he–him*) are highly reliable in transitive sentences (when they are present they indicate accurately agent and patient), but they are not always available (often there are only full nouns, which are not case marked). Word order is almost always available in German, but it is often not reliable (because sometimes transitive sentences have the agent after the verb and the patient before it—which works because they are both case marked for role). These two dimensions of cues—availability and reliability—can be combined to give an overall measure of cue validity (Kempe & MacWhinney, 1998).

With respect to acquisition, Bates and MacWhinney (1987) predicted that children should acquire first those cues with highest cue validity. In addition, because sometimes several cues may indicate the same function—providing extra information—children should find it especially easy to comprehend prototypical transitive sentences with both word order and case marking (and perhaps other cues) working in coalition: the coalitions-as-prototypes model (Bates & MacWhinney, 1987). This should be true especially if, as is often the case, the prototype occurs very frequently. An agent of a transitive action, for instance, should be identified most easily by a German child if it is not only marked by its position before the verb but also by the relevant case marker. In a study of English- and Italian-speaking children, Bates et al. (1984) provided evidence for this approach by comparing the use of word order and animacy cues (agents tend to be animate, patients inanimate) in transitive sentences. They found that the high cue validity of word order in English led English 2-year-olds to rely on word order and ignore animacy when these two cues conflicted (i.e., when they heard "The pencil is kicking the cow" they tried to make the pencil kick), whereas the low cue validity of word order in Italian led Italian 2-year-olds to rely on animacy and ignore word order (making the cow kick the pencil).

Some researchers have proposed that the particular aspects of cue validity that children follow change over development. In a study with Hebrew-speaking children and adults, Sokolov (1988) found that cue availability—how often a cue occurs—played a stronger role in sentence interpretation for younger children, whereas cue reliability—the proportion of relevant sentences for which a particular cue correctly indicated agent or patient—played a stronger role for older children and adults. Of special importance, children have to notice which cue adults follow when two cues conflict (conflict validity). In many cases, this may be a quite drawn-out process, as the relevant conflict situations are sometimes fairly rare in the language children experience (McDonald, 1986). Supporting this general view, Matessa and Anderson (2000) found that in adult artificial language learning cue validity predicted which cues are used early in the learning process and conflict validity predicted which cues are used in later learning.

Cue cost (essentially, complexity) has been much less studied. Building on Slobin's (1982) local cues hypothesis, one claim is that "local cues" such as case marking can be processed on the spot without taking the entire sentence into account, whereas "distributed cues" such as word order impose a greater burden on short-term processing capacity (because sentential fragments need to be held in memory until the next relevant component is processed). Support for this hypothesis was provided by Lindner (2003), who found that early in development German children tended to rely on "local cues" and only later on "distributed cues," although Lindner's analyses did not involve a direct comparison between conditions in which the different cues supported or conflicted with one another. Studies that have made such a direct comparison have found that German preschool children comprehend sentences in which case marking and word order conflict, more poorly than sentences in which case marking and word order collaborate (e.g., Mills, 1977; Primus & Lindner, 1994; Schaner-Wolles, 1989). However, these studies differ as to the age at which German children accurately comprehend sentences with conflicting cues, probably because the different studies used only sentences with highly familiar verbs, and exactly which verbs were used varied between studies. The use of familiar verbs makes it possible that children could respond on the basis of only verb-specific knowledge (e.g., knowing only that the hitter comes before *hit*), whereas mature grammatical knowledge is based on verb-general, abstract knowledge of grammatical constructions.

In the current study, we investigated German children's understanding of word order and case marking cues in transitive sentences, and—unlike previous studies in the competition model framework—we did this using novel verbs. Our specific question was when German children come to understand that in their language case marking is a 100% reliable cue (even if it is not always available), whereas word order is not (even though it is quite often available). In our two experimental studies, we gave children test sentences that contained various combinations of word order and case marking cues—all grammatically correct, with animacy neutralized across agent and patient. In one condition, the two cues supported one another: Case marking and word order both indicated the first noun as the agent. In a second condition, these two cues were in conflict: Word order indicated the first noun as the agent, whereas case marking indicated the second noun as agent. Finally, in a third condition, agent and patient were case-marked ambiguously and therefore the only cue children could rely on was word order. Following Bates and MacWhinney's (1987) concept of coalitions-as-prototypes, we predicted that sentences containing multiple, redundant cues (as in the first condition) should be easiest to acquire. From McDonald's (1986) findings, we predicted that sentences containing conflicting cues (as in the second condition) should be the most difficult because robust knowledge of relative cue reliabilities—from relatively rare conflict situations—is needed for adult-like comprehension. The findings from this study should be relevant not only for elucidating basic processes of language development but also for elucidating processes of children's learning more generally because it addresses such domain-general issues as the role of prototypes, the individuation of particular cues from prototypes, and children's sensitivity to more local versus more distributed cues in sequential learning.

Study 1

As a preliminary study to our two comprehension experiments, we first looked at how German adults use word order and case marking in transitive sentences addressed to young children. Because the competition model predicts that the cue validity of word order and case marking should play a key role in children's comprehension—and that cue availability and reliability might play different roles at different points in development—we computed all values for these two cues from a corpus of child-directed speech.

The German grammar relevant to the current studies is as follows. In active transitive sentences, the agent of the action is subject and is marked with nominative case marking, and the patient is direct object and is marked with accusative case marking. For both of these, the case marking is either a special form of pronoun or a noun with a special form of determiner (e.g., *a* or *the*). For example, if a dog is agent, the form is *der Hund* (the + nominative dog) or *er* (he), whereas if a dog is patient, the form is *den Hund* (the + accusative dog) or *ihn* (him). Additional complexity comes from the fact that nominative and accusative marking take different forms when applied to nouns of different genders, and in some cases, they are not distinct. For example, unlike the example of *dog* above (which is masculine), if a cat is the agent, the form is *die Katze* (the + nominative cat), but if a cat is the patient, the form is exactly the same *die Katze* (the + accusative cat). This means that in some instances, case marking is not an available cue in the sense that it does not identify case role unambiguously. Finally, although in German transitive sentences agents typically come before the verb and patients after the verb, as in English, to highlight the patient pragmatically the reverse order may be used—with the case roles marked by case marking and unaffected by the reverse order. "Den **Hund** beisst der Mann" has the first noun, *Hund*, marked as accusative, and the second noun, *Mann*, marked as nominative and so, despite word order, the man is biting the dog.

Method

For our analysis, we used Child Language Data Exchange System (CHILDES) data of spontaneous speech by German mothers to six monolingual normally developing children (Szagun, 2004). At the time of the first recording, the children were 1.8 years old, and at the time of second recording, they were 2.5 years old. We analyzed the sample of 7,032 utterances previously examined by Stoll, Abbot-Smith, and Lieven (in press), which these authors had coded into syntactic construction types. We examined the following categories where transitives might occur—transitives, complex sentences, subject-predicate-other, and verb fragments—and extracted transitive verbs by hand. Sentences with transitive verbs were excluded when they involved idioms, such as *Hunger haben* "to be hungry" (literally: have hunger) and passive constructions were also excluded. First, we divided all transitive sentences into sentences with verbs that were highly causative with a volitional agent and affected patient, such as *schubsen* "to push," and those with verbs that did not have any causative meaning, such as *sehen* "to see" (Hopper & Thompson, 1980). Then, we analyzed whether the sentence was complete, that is, with two noun phrases or whether it was a fragment, that is, subject or object was dropped.

We coded all transitive sentences for case marking in terms of whether they were unambiguous (i.e., it was clear which noun phrase was agent and which was patient) or ambiguous and for word order (subject-first or object-first). In addition, we coded which kind of case marker (i.e., which lexical form) was used. We followed Kempe and MacWhinney's (1998) formula for calculating cue availability, cue reliability, and cue validity for the case marking and word order cues that assign agent and patient. Availability of a cue was therefore defined as the number of sentences in which a cue is present, divided by the total number of transitive sentences. Reliability of a cue was defined as the ratio of sentences in which a cue correctly indicated the agent, divided by the number of sentences in which the cue was present. Finally, cue validity was defined as the product of availability and reliability. For our main analyses, we included only transitive sentences with highly causative verbs because only they contain both agent and patient. Nevertheless, we also compared these with our data for the noncausative transitive sentences and report the differences. All coding was carried out by the first author, and an additional coder coded 15% of all sentences for reliabilities. There was a high level of agreement between coders (Cohen's Kappa = 0.92).

Results and Discussion

Of our final sample of 745 transitive sentences, 410 (55%) contained highly causative verbs, and 335 (45%) were without causative meaning. Of the 410 transitive sentences with highly causative verbs, unambiguous case marking was found in 351 transitive sentences (86%) and 59 sentences (14%) contained ambiguous case marking. Fifty-five (13%) of the highly causative transitive sentences were fragments; that is, they involved either subject or object ellipsis. Of all complete highly causative sentences (with two noun phrases), 280 (79%) had a subject-first word order and 75 (21%) had an object-first word order.

To summarize, we found that in most (68%) of the complete causative transitive sentences (fragments excluded) both case marking and word order referred to the first noun as the agent. In 21% of the sentences, case marking and word order conflicted with each other because the second noun of the sentence was marked with nominative and/or the first one with accusative. In 11% of the sentences, word order was the only cue that referred to the first noun as the agent because the sentence contained ambiguous case marking. Only two sentences (less than 1%) appeared with an object-first order and ambiguous case marking.

When transitive sentences with noncausative verbs were included, similar results were found apart from the fact that object-first order appeared a bit more often (33%) and subject-first order with unambiguous case a bit less (56%). The proportion of sentences with subject-first order and ambiguous case marking was identical (11%). Of special relevance to the experiments of Study 2, we should note that marking case in German by definite determiners is not the most common way of marking (especially) agents of transitive sentences (with causative verbs) because quite often pronouns, which are always case marked, are used. The use of definite determiners to mark agents or patients was quite rare for agents (16%) but frequent for patients (60%). Within the transitive sentences with causative verbs, 8% of the agents were marked with *der* (the + masculine + nominative), 7.5% were marked by *die* (the + feminine), and finally only 0.5% were marked with *das* (the + neuter). Patients were mostly marked with *das* (30% of all transitive sentences), followed by *die* with 17%, and the rarest was *den* (the + masculine + accusative) at 13%.

Because we were interested in the relative strength of word order and case marking as cues for identifying agents, we calculated cue availability, cue reliability, and cue validity for both following Kempe and MacWhinney (1998). But whereas it is relatively easy to determine whether the case marking cue is available or not (unambiguous nominative and accusative forms), it is difficult to know exactly how German children use the word order cue. There are two possible ways. First, the position of one argument in relation to the verb might be sufficient to decide whether this noun phrase is agent or patient (subject-verb vs. verb-object). That is, *die Frau schubst* (the + feminine woman pushes) is likely to mean "the woman is pushing," whereas *schubst die Frau* (pushes the + feminine woman) is likely to mean "is pushing the woman." Under this analysis, the word order cue would also be available in fragment sentences (with either the subject or object omitted). Then, we find that the word order cue is available 100% of the time and the case-marking cue in 89% of the transitive sentences. In terms of reliability, however, case marking in German, when available, always reliably indicates the agent and patient of a transitive sentence 100% of the time, whereas we find that word order does this reliably only 74% of the time (because objects can come before, and subjects after, the verb). Therefore, the cue validity for case marking is higher with 89% compared to 74% cue validity for word order.

However, there is a second possible way to calculate the availability of word order because in German the position of the verb in the sentence is relatively flexible. It can either be at the beginning of a sentence as in questions, in the middle as in main clauses, or at end in subordinate clauses. Therefore, in a sentence such as, "..., *weil der Mann den Jungen schubst*" (..., because the + masculine + nominative man the + masculine + accusative boy pushes) the object (patient) comes directly before the verb, although the word order still maintains the most common (canonical) subject before object order. Thus, in fragment sentences without case marking, it is very difficult to say whether a noun phrase immediately before the verb is the agent or the patient ("*hat die Frau geschubst*" could either mean "he has pushed the woman" or "the women pushed him"). If we therefore decide that the word order cue is not available in German fragment sentences, that is, those with subject or object omission because the child needs to hear the relation between two arguments in the sentence to use the word order cue, we find even stronger differences. In this case, the availability of word order drops (87%) to almost the same as that of case marking (86%) and, once validity is calculated (with 100% reliability for case marking and 79% reliability for word order), case marking is even more valid with 86% cue validity in contrast to only 68% cue validity for word order (see Figure 1). This may be the more accurate calculation because German children hear transitives with subject or object omission 13% of the time. However, we will compare our data from the following experiments with both calculations using the different definitions of word order cues.

Study 2

We use the findings from Study 1 to make various predictions about which kinds of transitive sentences German children should comprehend most readily and at the earliest ages. If what is most important from the beginning is cue reliability—as suggested by MacWhinney et al. (1984)—or cue cost—as suggested by the local cues hypothesis (Slobin, 1982)—then children should comprehend most readily sentences with unambiguous case marking regardless of the order in which the noun phrases occur (i.e., even in object-first sentences). On the other hand, if what is most important from the beginning is cue availability—based mainly on frequency in the input—then they should comprehend most readily sentences in

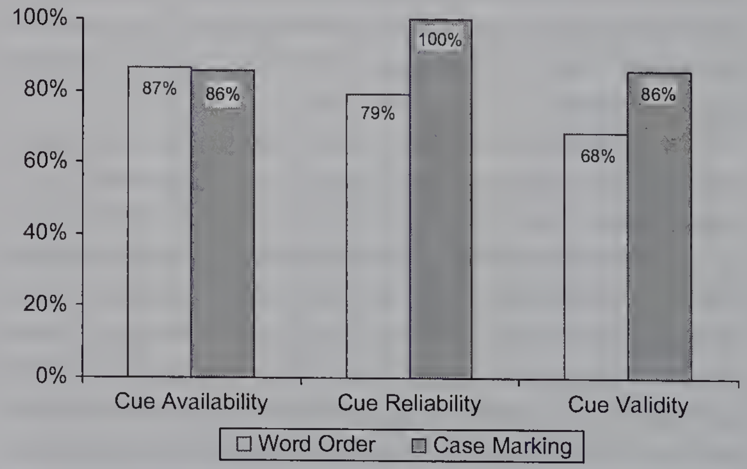


Figure 1. Availability, reliability, and validity of the grammatical cues word order and case marking for German transitive sentences in child-directed speech (Study 1).

which the agent is the first noun phrase, regardless of case marking (i.e., even in sentences with ambiguous case marking). Finally, if prototype sentences with redundant marking have a special role—as suggested by the coalitions-as-prototypes approach of Bates and MacWhinney (1987)—then children should comprehend most readily prototype sentences and might be expected to struggle when the cues conflict (i.e., in object-first sentences). Of course, it is also possible as we pointed out in the introduction and as suggested by Sokolov (1988) that cue availability, cue reliability, and prototypes play different roles at different periods of development.

In Study 2, we test these predictions experimentally using an act-out comprehension task, which is the task used most often in previous investigations of the competition model and local cues hypothesis. We adapted this task to examine how young German children perform when they hear sentences containing **novel** verbs to determine when and in which developmental order they start to use these grammatical cues productively, independent of any particular known verbs and independently of animacy cues.

Method

Participants

Sixteen monolingual German 2.7-year-old children (range = 2.6–2.8; 9 girls and 7 boys) and 16 monolingual German 4.10-year-old children (range = 4.6–5.3; 9 girls and 7 boys) were included in the study. An additional 9 children were tested but excluded from the study due to either fussiness (3), bilingualism (1), experimenter error (4), or because the child was too young (1). Children were recruited from a database of parents who volunteered to participate in psychological studies. They came from diverse social economic backgrounds. All children were tested in nursery schools in a medium-sized German city.

Materials

All verbs referred to prototypical causative-transitive actions, involving direct contact between a volitional agent and an affected patient. All actions were reversible (Hopper & Thompson, 1980). The two novel verbs *wiefen* and *tammen* were used to describe two novel transitive actions that were performed with two novel apparatuses. *Wiefen* was used to refer to an animal rocking another animal standing on a rocking-chair-like apparatus by pulling it toward itself with its head. *Tammen* referred to an animal pushing down another animal standing on a platform with a spring underneath it by jumping on its back.

Agents and patients of a particular event were pairs of animals with the same grammatical gender. Exactly which gender depended on the condition. The animals were all well known to 2-year-olds. We used the Elternfragebogen (Grimm & Doil, 2001), a much-shortened German version of the MacArthur Communicative Development Inventory (Fenson et al., 1994) to identify which animals to use. *Der Hase* "the (+masculine) bunny," *der Bär* "the (+masculine) bear," *der Elefant* "the (+masculine) elephant," *der Hund* "the (+masculine) dog," *die Katze* "the (+feminine) cat," and *das Schwein* "the (+neuter) pig" were on the Elternfragebogen. *Der Löwe* "the (+masculine) lion," *der Frosch* "the (+masculine) frog," and *der Tiger* "the (+masculine) tiger" were on the U.S. American MacArthur. Just two animals (*das Zebra* "the [+neuter] zebra" and *die Ziege* "the [+feminine] goat") were on neither of them, but the children did not show any difficulty in identifying these animals (see the Procedure section).

All children heard the same test sentences (see Appendix A) in three conditions: In the prototype condition, they heard the novel verbs with an argument structure in which the agent was the first noun phrase and case marked with nominative, and the patient was the second noun phrase and case marked with accusative, for example, *Der Hund wieft den Löwen.* "The (+nominative) dog is weefing the (+accusative) lion." In the "word-order-only" condition, they heard an argument structure in which the agent was the first noun phrase and the patient was the second noun phrase, but case marking was ambiguous because animals of masculine gender were not used, for example, *Die Katze wieft die Ziege.* "The cat is weefing the goat." In the conflict condition, the patient was the first noun phrase and case marked with accusative and the agent was the second phrase and case marked with nominative, for example, *Den Bären wieft der Tiger.* "The (+accusative) bear is weefing the (+nominative) tiger" (with the meaning: it is the tiger that is weefing the bear). As a control condition, we used one familiar verb *schubsen* "to push" in the prototype argument structure, for example, *Der Hund schubst den Tiger.* "The (+nominative) dog is pushing the (+accusative) tiger." Therefore, each child heard seven test sentences, six with novel verbs and one with a familiar verb.

Design

We tested each child with the two different novel verbs and the familiar verb in transitive sentence structures using an act-out task. A camera in front of the children recorded their enactment. Counterbalancing was used for the agent (e.g., lion/dog) and for sides, for example, sometimes the agent was to the left and sometimes to the right of the patient. The order of the verbs and the conditions was counterbalanced by Latin squares. There were 72 possible orderings of which 16 were chosen randomly, and these were distributed evenly over the children within each age group.

Procedure

During the session, the child sat at a small children's table on which the apparatuses for the act-out task were placed. The experimenter sat next to the child. Animals and apparatuses for the act-out task were hidden in a box. The two animals for each act-out task were always placed by the experimenter in front of the child between the child and the apparatus facing the child so that it was never the case that one animal was nearer to the apparatus. We counterbalanced which animal (agent or patient) was to the left of the child both within and between subjects.

Warm-up. The children first experienced a warm-up in which they were required to imitate acting out an intransitive locative (namely: *Der Fisch springt über den Elefanten.* "The [+nominative] fish is jumping over the [+accusative] elephant"). If they did not correctly act this out, they got a second trial with the sentence *Der Fisch klettert auf den Elefanten.* "The (+nominative) fish is climbing onto the (+accusative) elephant." If the child passed one warm-up trial correctly we proceeded with the experiment.

Word learning training. Prior to all three-test conditions each child was taught the name of the novel verb in the following manner. Using animals which take German feminine gender that does not decline for nominative or accusative case (e.g., *Kuh* "cow" and *Ente* "duck"), every verb (novel and familiar) was presented to each child in a live act out by the experimenter in a variety of argument structures: in the citation form with no arguments (e.g., *Das heißt wiefen.* "That's called weefing") as well as in transitive argument structure with two feminine pronouns (which are identical for subject and object position in German) in three different tenses (*Sie wird sie wiefen.* "She is going to weef her"; *Sie wieft sie.* "She is weefing her"; *Sie hat sie gewieft.* "She weefed her"). The child was also asked to repeat the verb in the citation form (e.g., *Kannst du das sagen: wiefen?* "Can you say this: weefing?") and to attempt the act out with the two feminine animals.

Test trial. For the act-out trials, the experimenter placed two animals in front of the child and told the child the test sentence: *Jetzt bist du dran! Zeig mir: Der Löwe wieft den Hund.* "Now it's your turn! Show me: The (+nominative) lion is weefing the (+accusative) dog." The experimenter repeated the test sentence until the child started enacting.

Vocabulary/morphology production posttest. After all test trials were over, the children took part in a language development test. The 2.7-year-olds received the vocabulary production subtest of the SETK 2 that has been standardized for German 2- to 3-year-olds (Grimm, 2000). In this, the children are shown cards with pictures of objects that they have to name. The 4.10-year-olds received the morphological production subtest of the SETK 3–5, which has been standardized for German 3- to 5-year-olds (Grimm, 2001). In this, children are shown pictures with familiar and novel objects and they had to build the correct plural form (of which there are eight possibilities in German). The 2.7-year-old children who participated in the test had a mean score of 44 (range = 36–56), and the 4.10-year-olds had a mean score of 47 (range = 36–63). The mean scores were, therefore, a bit lower than the expected ones for their age range (expected $M = 50$, $SD = 40-60$).

Coding and Reliability

For every test trial, the correct response was to choose the correct animal as agent of the action. If the child did not act out a causative scene but instead put both animals next to each other onto the apparatus we excluded those trials. We had to exclude 26 trials of 144 in the younger age group (prototypical condition [9], word-order-only condition [9], and conflict condition [8]) and none in the older age group. All children were coded by the first author, and an additional coder coded 15% of all trials for reliabilities with high agreement with the first author (Cohen's Kappa = 0.8774).

Results and Discussion

The data were analyzed using a 2 (age) × 4 (experimental condition) mixed factorial analysis of variance (ANOVA). There were main effects for both Condition, $F(3, 81) = 3.018$, $p < .05$, and Age, $F(1, 27) = 17.672$, $p < .001$, but not a significant Condition × Age interaction. Post hoc tests with a Bonferroni correction for the main effect of condition with six comparisons revealed only significant differences between the 4.10-year-old's performance with the familiar-verb control condition (M = 94%) and the conflict condition, $M = 56\%$, $t(15) = -4.392$, $p < .05$. Nonparametric tests (Wilcoxon) found the same result.

Because the chance level for our dependent variable was always 50%, we also investigated in which conditions and at which ages the children were above chance. The results show that the 2.7-year-olds were only above chance with the familiar verb, $t(15) = 2.236$, $p < .05$. In contrast, the 4.10-year-olds were above chance in the familiar-verb condition, $t(15) = 7.000$, $p < .001$, the prototypical, $t(15) = 3.576$, $p < .05$, and the word-order-only condition, $t(15) = 3.478$, $p < .05$, see Figure 2. However, we did not find any correlation between the children's performance in this task and vocabulary/morphology scores and also no group differences when comparing high and low vocabulary/morphology children.

Thus, 2.7-year-old German children were only able to comprehend transitive sentences in this act-out task with a familiar verb. One possible reason for this is that children initially form grammatical schemas around familiar verbs and are therefore only able to comprehend transitives sentences correctly with familiar verbs (Tomasello, 2003). However, a second explanation of the results is that the act-out task is particularly difficult for young children and it might be easier to carry out when asked to perform a known action than a novel action. By contrast, the German 4.10-year-olds correctly interpreted transitive sentences with novel verbs in subject-first word order, presumably because they have productive knowledge of the grammatical cue word order. But in the conflict condition, they performed at chance level. This indicates that German 4.10-year-olds have not yet acquired the use of the case-marking cue separately from subject-first word order and therefore do not interpret correctly object-first sentences.

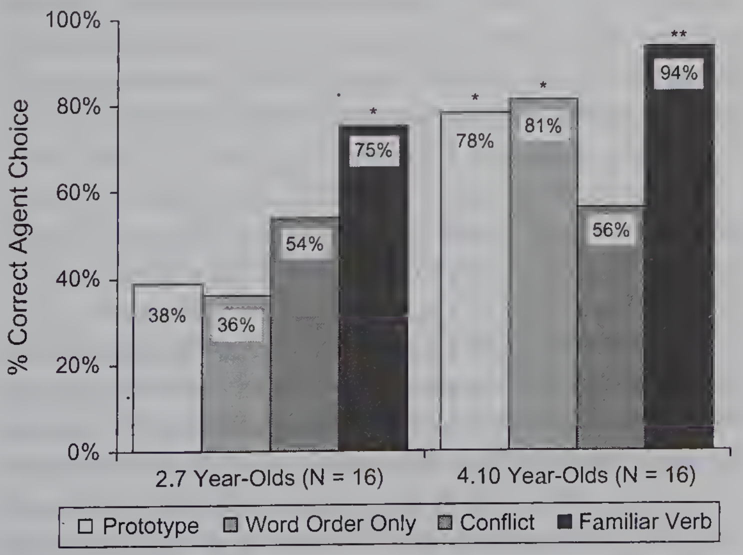


Figure 2. Mean proportion of correct agent and patient choices in the act-out task of Study 2, as a function of age and sentence type.

Study 3

It might be argued that the reason we found such late acquisition of case marking and verb-specific behavior in Study 2 is that the act-out task we used has high working memory and executive function demands. Some support for such an argument might be drawn from a previous study by Dittmar, Abbot-Smith, Lieven, and Tomasello (2008), who—contrary to the findings of the current experiment—found that German-speaking 2.6-year-olds did show productivity with novel verbs in transitive sentences in a pointing comprehension task. Therefore, in the next experiment, we adapted the pointing task to examine relative reliance on word order and case marking, using the same three novel verb conditions we used in Study 2. Furthermore, we tested a third age group of older children to try to identify a later point in language development when German children are able to comprehend object-first transitive sentences.

Method

Participants

The children were monolingual speakers of German, who were brought by a caregiver to a child lab in a medium-sized German city. Of these sixteen 2.7-year-old children (range = 2.6–2.8; 8 girls and 8 boys), sixteen 4.10-year-old children (range = 4.6–5.2; 8 girls and 8 boys), and sixteen 7.3-year-old children (range = 7.0–7.11; 8 girls and 8 boys) were included in the study. An additional 13 children were tested but excluded from the study due to either showing a side bias during the test trials (2), fussiness (7), bilingualism (2), or experimenter error (2). Children were recruited from a database of parents who volunteered to participate in psychological studies. They came from diverse social economic backgrounds.

Materials

All novel verbs referred to prototypical causative-transitive actions, involving direct contact between a volitional agent and an affected patient. Actions were reversible and involved either a caused change-of-state or change-of-location (Hopper & Thompson, 1980). The three novel verbs *wiefen*, *tammen,* and *baffen* were used to describe three novel transitive actions that were performed with three novel apparatuses. *Wiefen* was identical with the action used in Study 2 except that we emphasized the causality of this new event by making the agent force the patient into a handstand with the third repetition of the action. *Tammen* was also identical with the action used in Study 2 but we emphasized the causality of this new event by making the agent force the patient to fall sideways with the third repetition of the action. The third novel verb, *baffen,* referred to an animal spinning around another animal standing on a disk, and, with the third repetition of the action, the location of the patient was changed from being next to the agent to being further away.

Agents and patients of a presented event were the same pairs of animals as in Study 2 plus three more: *das Schaf* "the (+neuter) sheep" and *das Pferd* "the (+neuter) horse," which were on Elternfragebogen and *der Affe* "the (+masculine) monkey" which was on the U.S. American MacArthur-Bates Communicative Development Inventory (Fenson et al., 1994). The structural pattern of the test sentences (see Appendix B) was the same as described in Study 2. Each of the three conditions was tested with each of the three novel verbs, so that the children got nine test sentences. Unlike in Study 2, we did not test familiar verbs.

Design

We tested each child with three different novel verbs in transitive sentence structures using a pointing task. During the session, the children sat in front of a 31 × 49 cm "Apple Cinema Display" screen. The procedure of the pointing task was based on the Intermodal Preferential Looking (IPL) paradigm pioneered by Golinkoff, Hirsh-Pasek, Cauley, & Gordon (1987; see also Imai, Haryu, & Okada, 2005 for an adaptation of the IPL to a pointing task). For the test trials, the child saw two film scenes on the computer screen, each starting simultaneously and lasting 6 s. Both involved animals enacting the same causative event and differed only in that agent and patient roles were reversed. All children got alternating test sentences with the three different conditions and all three novel verbs were tested in one session.

For each test trial scene pair we counterbalanced which particular scene correctly matched the test sentence (e.g., for the pair "dog weef lion" and "lion weef dog" half of the children heard the German equivalent of "the dog is weefing the lion" and the other half heard the reverse). The order of the verbs and the conditions was counterbalanced by Latin squares. The target screen order was counterbalanced so that each side (left [L] or right [R]) was correct four or five times out of nine trials for each child (depended on counterbalancing order). The same side was never the correct choice more than twice in a row. No child experienced a condition in which the correct choice alternated regularly (e.g., LRLRLRLRL). For half of the children, the first correct side in the first trial was left and vice versa. There were 52 possible orderings for correct side of which 16 were chosen randomly and these were distributed evenly over the children within each age group. The direction of the action (from left to right or from right to left) was also counterbalanced.

Procedure

One camera from behind the children recorded their pointing behavior. Only children of the youngest age group sat on their parents' lap. When testing the older children, the parent sat behind the child on a separate chair. The parents whose children sat on their laps were asked to close their eyes during each test trial so as not to influence their child during pointing. We decided not to use headphones for the parents because we found that this distracted the children. The experimenter herself never looked at the screen during the test trials but always at the child.

Pointing practice training. To teach the children that the aim of the task was to point to one of two pictures on the computer screen, we used a very easy warm-up task with two pictures of objects, for example, "dog" and "duck," which appeared at the screen simultaneously. Then, the children were asked to point to one of the two objects (e.g., *Zeig mir das Bild: Das ist der Hund.* "Show me the picture: That's the dog"). The pictures were from the vocabulary comprehension subtest of the SETK-2 (Grimm, 2000). We repeated this task 10 times with different objects and all children solved it perfectly.

Word learning training. Similar to Study 2, every novel verb was presented to each child in a live act out by the experimenter in a variety of argument structures.

Film familiarization trials. Following the live enactment, for each verb the child then saw a familiarization trial in which he or she watched each of the two film scenes individually and heard the experimenter describing them in the citation form, for example, *Guck mal, das heißt wiefen.* "Look, that's called weefing," whereas the other half of the screen remained blank. The side where the children saw the first picture (left or right) was counterbalanced across and within subjects. At the end of each film scene, the experimenter pointed to each animal and asked the child "Wer ist das?" (Who's that?). The majority of the children had no problem spontaneously naming the participating animals. If a child did not name one of the animals, the experimenter told the child the name and asked him or her to repeat it, which almost all children then did.

Test trial. Following this a red dot focused the child's attention to the center of the computer screen. Then, the test trial began and the child watched the same two scenes as in the familiarization trials. But here they appeared simultaneously and were accompanied by a prerecorded linguistic stimulus with the target verb in transitive argument structure, for example, *Guck mal, der Löwe wieft den Hund.* (×2) "Look, the (+nominative) lion is weefing the (+accusative) dog." After the videos had stopped, the experimenter asked the child to point to the correct (still) picture by asking, for example, *Zeig mir das Bild: Der Löwe hat den Hund gewieft!* "Show me the picture: The (+nominative) lion weefed the (+accusative) dog!" If the child did not point the experimenter repeated the question a second time, but she never asked the child to point again once she/he had already done so.

Vocabulary/morphology production posttest. After all test trials were over, the children took part in a language development test. The 2.7-year-olds and the 4.10-year-olds received the same tests as in Study 2. The 7.3-year-olds received the morphological production subtest of the Heidelberger Sprachentwicklungstest in which children are shown pictures with familiar and novel objects and they had to form the correct plural or singular. This test has been standardized for 3- to 9-year-old Germans (Grimm & Schöler, 1998). The 2.7-year-old children achieved a mean score of 55 (range = 42–71), the 4.10-year-olds achieved a mean score of 56 (range = 38–69), and the 7.3-year-olds achieved a mean score of 49 (range = 40–59). The expected mean score is again 50 with a standard deviation between 40 and 60.

Coding and Reliability

For every test trial, the correct response was to choose the right animal as agent of the action. If the child did not choose either scene or pointed to both we excluded those trials. We had to exclude 19 trials of 144 in the youngest age group (prototypical condition [4], word-order-only condition [5], and conflict condition [10]), 1 (conflict condition) in the 4.10-year-olds, and none in the oldest age group. All children were coded by the first author, and an additional coder coded 15% of all trials for reliabilities with high agreement with the first author (Cohen's Kappa = 0.968).

Results and Discussion

The pointing behavior was analyzed using a 3 (age) × 3 (experimental condition) mixed factorial ANOVA. There were main effects for both condition, $F(2, 90) = 34.875, p < .001$, and age, $F(1, 45) = 19.258$, $p < .001$. However, these must be interpreted in the context of a significant Condition × Age interaction, $F(4, 90) = 5.855, p < .001$ (see Figure 3).

Post hoc tests with a Bonferroni correction for three comparisons revealed that the interaction was due to the 2.7-year-olds showing more correct pointing in the prototypical condition (M = 77%) than in the word-order-only condition, M = 50%, $t(15) = 2.595$, $p = .06$, and than in the conflict condition, M = 46%, $t(15) = 3.143$, $p < .05$. No difference was found between the word-order-only condition and the conflict condition. The 4.10-year-olds also pointed correctly more often in the prototypical condition (M = 88%) than in the conflict condition, M = 35%, $t(15) = 4.970$, $p < .001$, and also more often in the word-order-only condition (M = 94%) than in the conflict condition, $t(15) = 6.586$, $p < .001$. No difference was found between the prototypical condition and the word-order-only condition. The pattern of results for the 7.3-year-olds was generally the same as for the 4.10-year-olds—except that their performance in the conflict condition was much better (though still lower than the other conditions). That is, they pointed correctly more often in the prototypical condition (M = 98%) than in the conflict condition (M = 69%), $t(15) = 3.416$, $p < .05$, and more often in the word-order-only condition (M = 100%) than in the conflict condition, $t(15) = 3.758$, $p < .05$,—with no difference

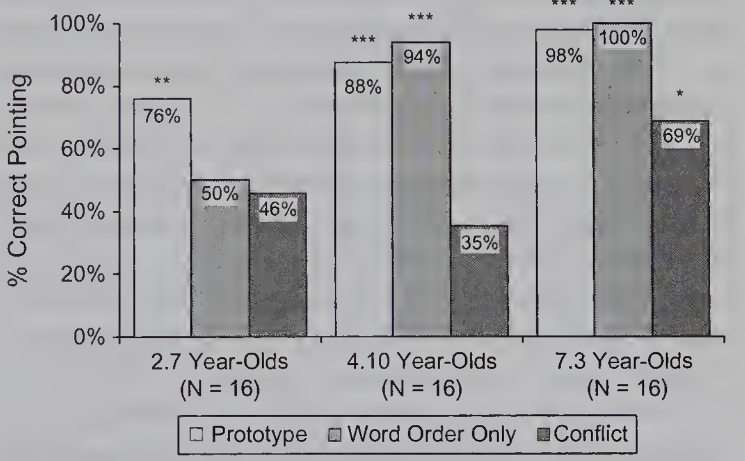


Figure 3. Mean proportion of pointing to the correct agent in Study 3, as a function of age and sentence type.

between the prototypical condition and the word-order-only condition.

Post hoc tests for the main effect of condition with Bonferroni correction revealed significant differences between all children's performance in the prototype condition ($M = 87\%$ correct pointing) and the conflict condition ($M = 50\%$ correct pointing, $t(47) = 6.601$, $p < .001$, and between the word-order-only condition ($M = 81\%$ correct pointing) and the conflict condition, $t(47) = 5.447$, $p < .001$. This indicates that conflicting cues, here word order and case marking, are especially difficult to use for children of all ages. Nonparametric tests (Wilcoxon) showed the same result.

Because the chance level for our dependent variable was always 50%, we also investigated in which conditions and at which ages the children were above chance. The results reflect the previous analyses, namely the 2.7-year-olds were only above chance in the prototypical condition, $t(15) = 4.354$, $p < .01$, whereas the 4.10-year-olds were above chance in both the prototypical, $t(15) = 9.121$, $p < .001$, and the word-order-only condition, $t(15) = 13.174$, $p < .001$, but not with the conflict condition. And finally the 7.3-year-olds reached ceiling in the prototypical and the word-order-only condition and were above chance in the conflict condition, $t(15) = 2.249$, $p < .05$. All analyses reflect a developmental trend whereby German children first acquire prototypical grammatical marking, followed by word order and only very late in age do they show an adult-like reliance on case marking when this conflicts with word order.

We were interested in what strategies young German children use to interpret transitive sentences with patients in first position. Therefore, we analyzed all children's responses to the conflicting sentences to see whether they oriented toward word order or case marking or whether they used neither strategy and avoided selecting a scene (usually through pointing to both scenes). A 3 (age) × 3 (strategy) mixed factorial ANOVA revealed main effects for both strategy, $F(2, 90) = 23.473$, $p < .001$, and age, $F(1, 45) = 21025.000$, $p < .001$. However, these must be interpreted in the context of a significant Strategy × Age interaction, $F(4, 90) = 6.362$, $p < .001$. Post hoc tests with a Bonferroni correction for three comparisons showed that 4.10-year-olds relied significantly more on word order than 7.3-year-olds, $t(30) = 2.622$, $p < .05$, and 7.3-year-olds relied more on case marking than 4.10-year-olds, $t(30) = -2.879$, $p < .05$, and 2.7-year-olds, $t(30) = -3.922$, $p < .001$ (see Figure 4).

Furthermore, we found that the performance of the 4.10-year-olds in the conflict condition was related to their state of morphological knowledge (plural morphology). Children who performed poorly on the morphological productivity posttest relied more strongly on word order in our experiment and therefore pointed incorrectly in the conflict condition ($M = 17\%$ correct pointing) than children with more robust morphological knowledge ($M = 54\%$ correct pointing), $t(14) = -2.460$, $p < .05$. The low-morphology group of children even showed below chance performance in the conflict condition, $t(7) = -5.372$, $p < .01$, which indicates a word order strategy. Similar findings come from the high-morphology group of 7.3-year-olds who showed above chance performance in the conflict condition, $t(7) = 3.122$, $p < .05$, whereas the low-morphology group of children still performed at chance. It may therefore be the case that German children pass through a stage in which they rely solely on word order and ignore case marking when these cues conflict before they learn to rely solely on case marking as adults do.

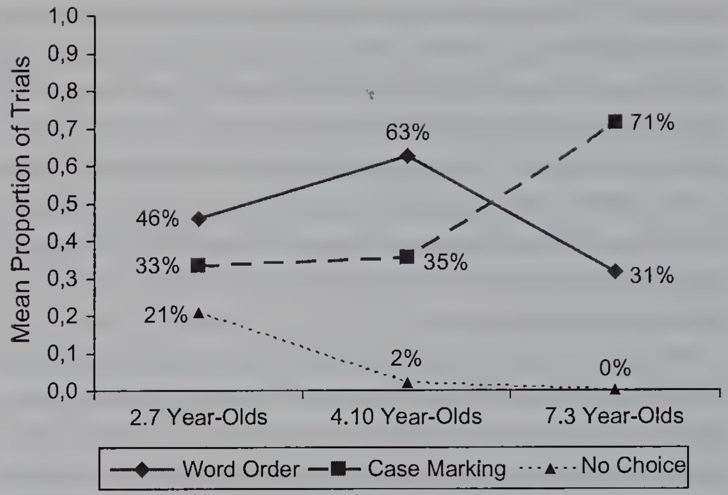


Figure 4. Strategies used (cue types relied on) during trials with conflicting cues in Study 3, as a function of age.

Our findings from Study 3 support the hypothesis that transitive sentences with a subject-first word order and with unambiguous case marking are acquired earlier by German children than are transitive sentences with a subject-first word order but ambiguous case marking. Furthermore, at age 5, German children have still problems correctly comprehending transitive sentences with object-first word order even when these are clearly case marked. By age 7, the majority of the children have solved this problem.

General Discussion

The current studies paint a fairly clear picture of how young German children come to comprehend causative transitive sentences. At around 2.6 years of age, if assessed with an act-out task (Study 2), they comprehend transitive sentences with familiar verbs but not novel verbs. This finding is in general agreement with the production study of Wittek and Tomasello (2005) in suggesting fairly verb-specific knowledge early in development. However, when a less demanding pointing task is used (Study 3), German children at this same age show solid comprehension of prototypical transitive sentences in which both word order and case marking indicate who was doing what to whom redundantly—even with novel verbs, suggesting more verb-general knowledge at 2.6 years. That they could show their knowledge only in the experiment using a pointing task and not in the act-out experiment might be due to the memory-burdensome nature of the act-out method per se with small children (Hirsh-Pasek & Golinkoff, 1996; Munakata, McClelland, Johnsons, & Siegler, 1997).

But, importantly, these children comprehended transitive sentences only in their prototypical form with redundant marking of agent and patient. Even with the less demanding pointing measure, they did not comprehend transitive sentences for which diagnostic case marking was absent or those in which the word order was noncanonical (object-first). They could not use either cue by itself, and they suffered when either was absent. These findings suggest that in languages like German children do not begin by attending to single cues, but rather they learn to comprehend the prototype and have difficulty whenever there is deviation from it. The prototypical form in German is also the most frequent (Study 1), presumably a common pattern crosslinguistically for case-marking languages. The role of subject–verb agreement in this process (and animacy as a semantic cue) should also be investigated.

The 4.10-year-old children present us with a puzzle. In both studies, using both methods, they seem to comprehend transitive sentences mostly in terms of word order. In both Studies 2 and 3, their performance with word order only is as high as with the full prototype including case marking (both near ceiling), and they choose at random in response to sentences in which word order and case-marking conflict—with a number of children in Study 3 actually ignoring case and going with word order only. This finding is a puzzle because on the two standard measures of input in the competition model—cue availability (how often the cue is available in relevant sentences) and cue reliability (how reliable the cue is, when it is present, in indicating the correct interpretation)—word order shows no advantage in availability (87% vs. 86% for case marking), and indeed its cue reliability as standardly computed is lower (79% vs. 100% for case marking).

One possible explanation of this finding is that the way we are thinking about grammatical cues is not fully adequate. It may be that cue availability and reliability as calculated here for word order miss aspects of the input that are important for language learning children. First, as noted above, it may be that German children do not use the word order cue as the positional relation between the two nouns in the sentence (first noun = agent; second noun = patient) but as the positional relation between the noun and the inflected verb (noun before verb = agent; noun after verb = patient). That would mean that the word order cue is also available in fragment sentences and hence more often available (100%) than case marking (89%). It is also possible that German children use the word order cue as the positional relation between the two nouns but do not take fragment sentences (with subject or object omission) as part of the transitive domain.

A second possibility, also alluded to above, is that German children do not use case marking in a completely general way. Because German has three noun classes, nominative case marking, for example, has three different forms in the singular and another in the plural. If children at a particular age have not yet discovered that all these forms mark the same case, then the way that cue reliability is typically calculated is not fully adequate. That is, the children in the current studies were tested on the particular case markers *der* and *den* used as determiners (masculine nominative and accusative), which appear in only 21% of all transitive sentences, and their comprehension of these may not benefit from their experience with case marking using pronouns, in which case the cue availability of these particular forms is not particularly high. But, of course, as children learn to connect the different case-equivalent forms (e.g., the nominative forms for nouns of different genders, as well as the nominative form for personal pronouns of the same gender), the cue availability of case marking will go up (even if the input stays exactly the same). Calculating the cue availability of case marking in this more item-based way results in the availability of case marking being much lower (21%) than that of word order (87%) even when assuming that word order is not available in fragment sentences.

Both approaches to calculating the cue availability of word order and case marking result in the conclusion that availability might indeed be higher for word order than for case marking. With this prediction, it would not be unexpected anymore for our 4.10-year-old children to rely more on word order than on case marking. This suggests that young German children rely on different input parameters at different stages of development; specifically, they rely more on cue availability (basically frequency) early in development and more on cue reliability later in development (see Sokolov, 1988, for similar findings). In agreement with this view, many studies have demonstrated the importance of frequency in early language development (see Lieven & Tomasello, 2008, for a review). Complicating matters further, many of the case markers in German are either not diagnostic within the transitive (*die* is both the nominative and accusative feminine; *das* is both the nominative and accusative neuter) or else ambiguous with forms outside the transitive (e.g., the masculine nominative form *der* is also the feminine dative and genitive). It is also important that in online sentence processing, German adults show faster reaction times when the test sentence only has a cue with high availability rather than one with high reliability (Kempe & MacWhinney, 1999)—even though in offline (less time-pressured) agent identification tasks they rely more on the cue with high reliability (see also the artificial language learning tasks of Matessa & Anderson, 2000). In all, it would seem that German word order is somehow a more straightforward cue for younger, less grammatically sophisticated children than is German case marking, which has so many different and ambiguous forms for the same grammatical function.

It must also be noted that the finding that German 4.10-year-olds rely more on word order than case marking does not accord well with Slobin's (1982) local cues hypothesis, which would predict the "local" case-marking cue to be easier to process than the "distributed" word order cue. However, German case marking differs in two ways from case marking in languages such as Turkish or Hungarian on which the local cues hypothesis was based (MacWhinney, Pleh, & Bates, 1985 for Hungarian; Slobin & Bever, 1982 for Turkish). First, whereas in Turkish and Hungarian case is marked by suffixes on the noun, in German case is marked on the determiner or adjective that precedes the noun. Therefore, one might suggest that case marking is not as local in German as in Turkish or Hungarian. Second, as just noted, the form of the German masculine nominative determiner *der* and accusative determiner *den* is ambiguous with determiner forms outside the domain of transitive sentences. Both factors, "less locality" and "ambiguity," may influence the ease of sentence or cue processing in German transitive sentences compared to Turkish or Hungarian.

Finally, we come to the 7.3-year-olds. We ourselves were very surprised that it was only at this late age that children succeeded in the conflict condition, weighing the case marking cue over the word order cue as adults do (Study 3). However, even adults have difficulties processing noncanonical word orders, at least as measured by reaction times (Ferreira, 2003; Kaiser & Trueswell, 2004). When German adults are confronted with object-first sentences that are ambiguously marked on the first noun phrase, they initially interpret these as subject-first sentences until they hear the second noun phrase (Weber, Grice, & Crocker, 2006). Moreover, in point of fact, our current findings do not differ greatly from those of other studies that have used familiar verbs. In the studies of Primus and Lindner (1994) and Schaner-Wolles (1989), it was not until children were 5 years of age that they correctly comprehended transitive sentences with familiar verbs with conflicting word order and case marking cues. Apparently, resolving conflicting cues in sentences with novel verbs takes even longer, though how much longer is not known as we did not test children between 5 and 7 years of age.

In terms of cue availability and reliability, following the reasoning from above, children by 7 years of age should know the grammatical equivalence of all (or at least most) of the different case forms serving the same grammatical function (and should ignore ambiguities based on other information). For 7-year-olds, then, the cue reliability of case marking is something close to that computed here; therefore, they finally rely on case marking over word order, as German adults would do. They also had much more experience than the younger children with sentences containing conflicting cues. One might argue that 21% object-first sentences in the input is plenty of exemplars for learning about conflicting cues by the age of 4.10; therefore, our children at this age should have been better. However, two other factors must be taken into account. First, object-first sentences occur in pragmatically marked contexts, with stress on the initial noun—which might mark them for children as a separate construction from prototypical transitive sentences without such stress. Second, almost all the object-first sentences in German child-directed speech have pronouns, not lexical nouns with determiners, in the pre- or postverbal position (96%), and most of these (76%) are first and second person personal pronouns with which the child is highly familiar. This means that the child can comprehend the vast majority of object-first transitive sentences on the basis of well-entrenched knowledge of specific pronoun forms and meaning but need not use case marking per se. Furthermore, the majority of the remaining 4% of the object-first sentences without pronouns provided an additional animacy cue to the child, that is, an animate agent versus an inanimate patient, despite the patient appearing in sentence-initial position. Overall, only 1% of all object-first sentences were based solely on the competition between the grammatical cues of case marking and word order. Therefore, in actual fact, young children hear very few conflict sentences in which they really are forced to decide between case marking and word order. This does not mean that the children in our experiments heard odd or ungrammatical sentences, just very infrequent ones if frequency is counted at the level of specific forms such as pronouns and particular case markers.

The overall process by which German children learn to comprehend transitive sentences in a verb-general way may therefore be summarized as follows. They begin somewhere after the second birthday by comprehending the prototypical form of such sentences (even with novel verbs) with redundant marking of agent and patient by means of word order and case marking. Between ages 2 and 4 they learn to use word order by itself, as well as a number of specific lexical forms like personal pronouns that appear in different case-marked versions. But it is only by sometime after age 5 that they become adult-like in weighing case marking over word order in sentences in which these cues conflict. Interestingly, this same process may help to explain why English-speaking children take so long to comprehend and produce sentences with novel verbs in experiments such as those summarized by Tomasello (2000). The prototypical transitive sentence in English potentially has animacy cues, a case marked subject pronoun (such as *I* or *he*), and subject–verb agreement—in addition to canonical subject-verb-object word order. In most of the experiments, all these cues were neutralized except word order. Following the reasoning of the current study, then, the prediction would be that English-speaking children should do better at an earlier age with prototypical transitive sentences including redundant cues. What this means is that all children learning all languages take time to learn the significance of individual cues when they experience those cues most often in combination with other redundant cues. This accords with the coalition model by Hirsh-Pasek and Golinkoff (1996) who suggested that children might master grammar by noting redundancies of cues for comprehension and with much recent theorizing in adult psycholinguistics in which the process of comprehension is seen as learning to integrate a great diversity of multiple probabilistic cues to language structure (the cue integration approach; see Christiansen & Monaghan, 2006).

In any case, the current study has demonstrated that even for what many researchers consider the most straightforward grammatical construction of all, the simple transitive, it can be a fairly long drawn-out process for young children to achieve adult-like mastery of the specific roles of each of the different grammatical cues instantiated in the particular sentences they hear. This mastery depends on their attention to basic aspects of their linguistic experience, such as the frequency, consistency, and complexity of those cues in particular utterances. Frequency, consistency, and complexity have also, of course, been centrally important in theories of children's nonlinguistic cognitive development and inductive learning (see Siegler, 1996). Our finding of the importance of the prototype and the long process of "unpacking" it into the different cues it contains, also finds resonance with much recent discussion of the relationship between prototypes and exemplar-based models in adult categorization learning (see, e.g., Anderson 1991; Chandler, 2002; Hampton, 1997; Ross & Makin, 1999). Although there is a great deal of theoretical and empirical work to be undertaken to make the links between these research fields more explicit, our results suggest to us that this aspect of children's language learning shows close parallels with essential characteristics of human learning more generally.

References

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review, 98*, 409–429.

Bates, E., & MacWhinney, B. (1987). Competition, variation, and language learning. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 157–193). Hillsdale, NJ: Lawrence Erlbaum.

Bates, E., & MacWhinney, B. (1989). Functionalism and the competition model. In B. MacWhinney & E. Bates (Eds.), *The cross-linguistic study of sentence processing* (pp. 3–73). New York: Cambridge University Press.

Bates, E., MacWhinney, B., Caselli, C., Devescovi, A., Natale, F., & Venza, V. (1984). A cross-linguistic study of the development of sentence interpretation strategies. *Child Development*, *55*, 341–354.

Chandler, S. (2002). Skousen's analogical approach as an exemplar-based model of categorization. In R. Skousen, D. Lonsdale, & D. B. Parkinson (Eds.), *Analogical modelling* (pp. 51–105). Philadelphia: John Benjamins.

Christiansen, M., & Monaghan, P. (2006). Discovering verbs through multiple-cue integration. In K. Hirsh-Pasek & R. Golinkoff (Eds.), *Action meets word. How children learn verbs* (pp. 88–107). New York: Oxford University Press.

Dittmar, M., Abbot-Smith, K., Lieven, E., & Tomasello, M. (2008). *Young children's comprehension of the transitive construction: A crosslinguistic study.* Manuscript submitted for publication.

Fenson, L., Dale, P., Reznick, J., Bates, E., Thal, D., & Pethnick, S. (1994). *Variability in early communicative development*. Chicago: University of Chicago Press.

Ferreira, F. (2003). The misinterpretation on noncanonical sentences. *Cognitive Psychology, 47*, 164–203.

Gertner, Y., Fisher, C., & Eisengart, J. (2006). Learning words and rules: Abstract knowledge of word order in early sentence comprehension. *Psychological Science, 17*, 684–691.

Goldberg, A. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K. M., & Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language, 14*, 23–45.

Grimm, H. (2000). *Sprachentwicklungstest für zweijährige Kinder. Diagnose rezeptiver und produktiver Sprachverarbeitungsfähigkeiten*. Göttingen, Germany: Hogrefe.

Grimm, H. (2001). *Sprachentwicklungstest für drei- bis fünfjährige Kinder. Diagnose von Sprachverarbeitungsfähigkeiten und auditiven Gedächtnisleistungen*. Göttingen, Germany: Hogrefe.

Grimm, H., & Doil, H. (2001). *Elternfragebogen für die Früherkennung von Risikokindern*. Göttingen, Germany: Hogrefe.

Grimm, H., & Schöler, H. (1998). *Heidelberger Sprachentwicklungstest*. Göttingen, Germany: Hogrefe.

Hampton, J. A. (1997). Psychological representation of concepts. In M. A. Conway (Ed.), *Cognitive models of memory* (pp. 81–110). Cambridge, MA: MIT Press.

Hirsh-Pasek, K., & Golinkoff, R. M. (1996). *The origins of grammar*. Cambridge, MA: MIT Press.

Hopper, P. J., & Thompson, S. A. (1980). Transitivity in grammar and discourse. *Language, 56*, 251–299.

Imai, M., Haryu, E., & Okada, H. (2005). Mapping novel nouns and verbs onto dynamic action events: Are verb meanings easier to learn than noun meanings for Japanese children. *Child Development, 76*, 340–355.

Kaiser, E., & Trueswell, J. C. (2004). The role of discourse context in the processing of a flexible word-order language. *Cognition, 94*, 113–147.

Kempe, V., & MacWhinney, B. (1998). The acquisition of case marking by adult learners of Russian and German. *Studies in Second Language Acquisition, 20*, 543–587.

Kempe, V., & MacWhinney, B. (1999). Processing of morphological and semantic cues in Russian and German. *Language and Cognitive Processes, 14*, 129–171.

Lieven, E., & Tomasello, M. (2008). Children's first language acquisition from a usage-based perspective. In P. Robinson & N. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 168–196). New York: Routledge.

Lindner, K. (2003). The development of sentence-interpretation strategies in monolingual German-learning children with and without specific language impairment. *Linguistics, 41*, 213–254.

MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German and Italian. *Journal of Verbal Learning and Verbal Behavior, 23*, 127–150.

MacWhinney, B., Pleh, C., & Bates, E. (1985). The development of sentence interpretation in Hungarian. *Cognitive Psychology, 17*, 178–209.

Matessa, M., & Anderson, J. R. (2000). Modelling focused learning in role assignment. *Language and Cognitive Processes, 15*, 263–292.

McDonald, J. L. (1986). The development of sentence comprehension strategies in English and Dutch. *Journal of Experimental Child Psychology, 41*, 317–335.

Mills, A. S. (1977). *First and second language acquisition in German. A parallel study*. Ludwigsburg, Germany: Ludwigsburg Studies in Language and Linguistics.

Munakata, Y., McClelland, J. L., Johnsons, M. H., & Siegler, R. S. (1997). Rethinking infant knowledge: Toward an adaptive process account of successes and failures in object permanence tasks. *Psychological Review, 104*, 686–713.

Primus, B., & Lindner, K. (1994). Variation in grammar and first language acquisition: A new concept of parameter in universal grammar. In R. Tracy & E. Lattey (Eds.), *How tolerant is universal grammar? Essays on language learnability and language variation* (pp. 185–204). Tübingen, Germany: Niemeyer.

Ross, B. H., & Makin, V. S. (1999). Prototype versus exemplar models. In R. J. Sternberg (Ed.), *The nature of cognition* (pp. 460–492). Cambridge, MA: MIT Press.

Schaner-Wolles, C. (1989). Strategies in acquiring grammatical relations in German: Word order or case marking? *Folia Linguistica, 23*, 131–156.

Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking*. New York: Oxford University Press.

Slobin, D. I. (1982). Universal and particular in the acquisition of language. In L. R. Gleitman & E. Wanner (Eds.), *Language acquisition: The state of the art* (pp. 128–170). Cambridge, UK: Cambridge University Press.

Slobin, D. I., & Bever, T. G. (1982). Children use canonical sentence schemas: A crosslinguistic study of word order and inflections. *Cognition, 12*, 229–265.

Sokolov, J. L. (1988). Cue validity in Hebrew sentence comprehension. *Journal of Child Language, 15*, 129–155.

Stoll, S., Abbot-Smith, K., & Lieven, E. (in press). Lexically restricted utterances in Russian, German and English child directed speech. *Cognitive Science*.

Szagun, G. (2004). Learning by ear: On the acquisition of case and gender marking by German-speaking children with normal hearing and cochlear implants. *Journal of Child Language, 31*, 1–30.

Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition, 74*, 209–253.

Tomasello, M. (2003). *Constructing a language*. Cambridge, MA: Harvard University Press.

Weber, A., Grice, M., & Crocker, M. W. (2006). The role of prosody in the interpretation of structural ambiguities: A study of anticipatory eye movements. *Cognition, 99*, B63–B72.

Wittek, A., & Tomasello, M. (2005). German-speaking children's productivity with syntactic constructions and case morphology: Local cues act locally. *First Language, 25*, 103–125.

Appendix A: Test Sentences Act-Out Experiment (Study 2)

(Half of the children heard the sentences with reversed agent and patient)

a. Prototype condition
Der Hund wieft den Löwen.
(The$_{\text{masculine·nomiative}}$ dog is weefing the$_{\text{masculine.accusative}}$ lion.)
Der Bär tammt den Elefanten.
(The$_{\text{masculine·nominative}}$ bear is tamming the$_{\text{masculine.accusative}}$ elephant.)
b. Word-order-only condition
Die Katze wieft die Ziege.
(The$_{\text{feminine}}$ cat is weefing the$_{\text{feminine}}$ goat.)
Das Schwein tammt das Zebra.
(The$_{\text{neuter}}$ pig is tamming the$_{\text{neuter}}$ zebra.)
c. Conflict condition
Den Tiger wieft der Bär.
(The$_{\text{masculine.accusative}}$ tiger is weefing the$_{\text{masculine.nominative}}$ bear.)
Den Hasen tammt der Frosch.
(The$_{\text{masculine.accusative}}$ bunny is tamming the$_{\text{masculine.nominative}}$ frog.)
d. Familiar verb condition
Der Tiger schubst den Hund.
(The$_{\text{masculine·nomiative}}$ tiger is pushing the$_{\text{masculine.accusative}}$ dog.)

Appendix B: Test Sentences Pointing Experiment (Study 3)

(Half of the children heard the sentences with reversed agent and patient)

e. Prototype condition
Der Hund wieft den Löwen.
(The$_{\text{masculine·nomiative}}$ dog is weefing the$_{\text{masculine.accusative}}$ lion.)
Der Bär tammt den Elefanten.
(The$_{\text{masculine·nominative}}$ bear is tamming the$_{\text{masculine.accusative}}$ elephant.)
Der Frosch bafft den Affen.
(The$_{\text{masculine·nominative}}$ frog is baffing the$_{\text{masculine.accusative}}$ monkey.)
f. Word-order-only condition
Die Katze wieft die Ziege.
(The$_{\text{feminine}}$ cat is weefing the$_{\text{feminine}}$ goat.)
Das Schwein tammt das Zebra.
(The$_{\text{neuter}}$ pig is tamming the$_{\text{neuter}}$ zebra.)
Das Schaf bafft das Pferd.
(The$_{\text{neuter}}$ sheep is baffing the$_{\text{neuter}}$ horse.)
g. Conflict condition
Den Tiger wieft der Bär.
(The$_{\text{masculine.accusative}}$ tiger is weefing the$_{\text{masculine.nominative}}$ bear.)
Den Hasen tammt der Frosch.
(The$_{\text{masculine.accusative}}$ bunny is tamming the$_{\text{masculine.nominative}}$ frog.)
Den Hund bafft der Elefant.
(The$_{\text{masculine.accusative}}$ dog is baffing the$_{\text{masculine.nominative}}$ elephant.)

Child Development, July/August 2008, Volume 79, Number 4, Pages 1168 – 1182

The Dynamics of Life Stressors and Depressive Symptoms in Early Adolescence: A Test of Six Theoretical Models

Margaret Clements and J. Lawrence Aber
New York University

Edward Seidman
William T. Grant Foundation

Structural equation modeling was used to compare 6 competing theoretically based psychosocial models of the longitudinal association between life stressors and depressive symptoms in a sample of early adolescents ($N = 907$; 40% Hispanic, 32% Black, and 19% White; mean age at Time 1 = 11.4 years). Only two models fit the data, both of which included paths modeling the effect of depressive symptoms on stressors recall: The mood-congruent cognitive bias model included only depressive symptoms to life stressors paths (DS → S), whereas the fully transactional model included paths representing both the DS → S and stressors to depressive symptoms (S → DS) effects. Social causation models and the stress generation model did not fit the data. Findings demonstrate the importance of accounting for mood-congruent cognitive bias in stressors – depressive symptoms investigations.

One theoretical approach to understanding the development of depression emphasizes the etiological role of stressful life circumstances. Early research in this area focused on major life stressors: acute and disruptive events that occur infrequently (Holmes & Rahe, 1967; Johnson, 1986). Later research incorporated minor stressors, events that occur more frequently, but are not as disruptive as major stressors (DeLongis, Folkman, & Lazarus, 1988; Kanner, Coyne, Schaefer, & Lazarus, 1981). School-related pressures or arguments are examples of minor stressors. Many cross-sectional and longitudinal studies have identified positive associations between both types of stressors and psychological symptoms, with most research inferring that experiencing multiple major and/or minor stressors leads to the development of psychological symptoms. However, there has been considerable debate regarding the validity of these conclusions.

One area of debate is whether and to what extent major and minor stressors actually lead to the development of depressive symptoms or whether a different causal process is in effect (Caspi, Elder, & Bem, 1987; Felner & Felner, 1989; Hammen, 1992b). The second debate focuses on the structure of the self-report measures typically employed in these studies. The concern here is that these measures include items that represent preexisting psychological adjustment, which results in inflated estimates of the association between life stressors and depressive symptoms (L. H. Cohen, Burt, & Bjorck, 1987; B. S. Dohrenwend, Dohrenwend, Dodson, & Shrout, 1984; Masten, Neeman, & Andenas, 1994). The current study contributes to the life stressors – depressive symptoms research by systematically comparing the fit of six competing theory-based psychosocial models, while at the same time addressing critiques regarding life stressors measurement.

Early Life Stressors Research: Three Social Causation Models

Most research investigating the association between stressful life circumstances and depressive symptoms applies a social causation theory of the stressors – depressive symptoms (S-DS) association, which suggests that stressors lead to symptoms. Early studies focused on major life stressors or "life events" (Coddington, 1972; Holmes & Rahe, 1967; Johnson, 1986). Although many cross-sectional and longitudinal studies found a positive relationship between major stressors and depressive symptoms in children and adolescents (e.g., Larson & Ham, 1993; Meyer et al., 1993; Towbes, Cohen, & Glyshaw, 1989), the effect sizes identified were typically small (Johnson, 1986; Rabkin

---

This research was supported in part by National Institute of Mental Health Grant MH43084 and Carnegie Corporation Grant B4850, awarded to Edward Seidman, J. Lawrence Aber, and LaRue Allen. We thank the children and schools whose cooperation made this study possible, as well as Patrick Shrout for his comments on an early draft of this manuscript.

Correspondence concerning this article should be addressed to Margaret Clements, Department of Applied Psychology, Steinhardt School of Culture, Education, and Human Development, New York University, 246 Greene Street, Floor 4R, New York, NY 10003. Electronic mail may be sent to peggy.clements@nyu.edu.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7904-0023

& Struening, 1976). In light of these small effect sizes, some researchers began to include other types of stressful circumstances. Lazarus and his colleagues emphasized the importance of the minor stressors that people encounter as part of daily life, emphasizing that these more proximal stressors may have a stronger association with maladaptive outcomes than major stressors (DeLongis et al., 1988; Kanner et al., 1981).

Previous research and theory have proposed three social causation models. Initial studies investigated an *independent effects* model, which suggests that both major and minor stressors have significant, independent effects on psychological adjustment (see Figure 1a). As Lazarus had anticipated, minor stressors explained more variance in psychological symptoms than did major stressors (Kanner et al., 1981), a finding that was replicated in cross-sectional studies of adolescents (DuBois, Felner, Meares, & Krier, 1994; Rowlison & Felner, 1988).

A second social causation model is a *mediated effects model*, which proposes that minor stressors result from the disruption caused by major stressors and mediate their impact (Rowlison & Felner, 1988; see Figure 1b. A study by Rowlison and Felner (1988) designed to test this theory found that minor stressors only partly mediate the association between major stressors and depressive symptoms and that both major and minor stressors have independent associations with psychological adjustment. In light of these findings, they suggested that a *partially mediated model* is more appropriate (see Figure 1c). However, findings have been inconsistent, with other adolescent researchers finding evidence of full mediation (Compas, Howell, Phares, Williams, & Ledoux, 1989; Wagner, Compas, & Howell, 1988). Further research is necessary to identify the process involved.

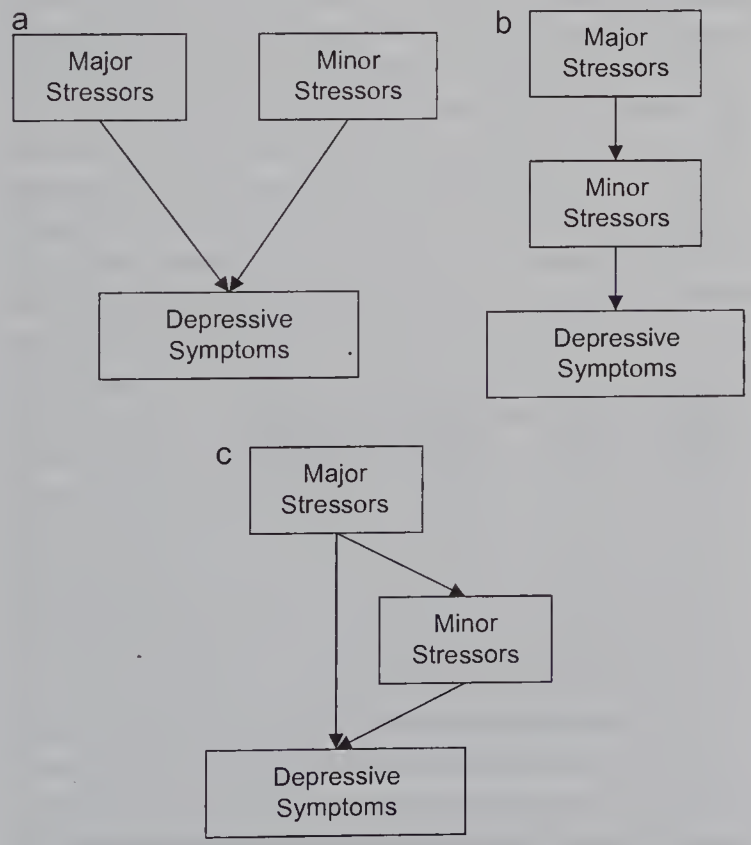


Figure 1. Heuristic models of social causation theories.

An important aspect of most S-DS studies is that they use a single wave of life stressors data to predict either a concurrent or a change in depressive symptoms. This research design continues to be used even though it has been subjected to two critiques. One critique focuses on the typical methods used to measure life stressors, and the second challenges the causal assumptions often made about the S-DS association. We review each of these issues below.

Life Stressors Measurement

One critique focuses on life stressors measurement techniques. Most studies use self-report checklist measures in which respondents check off the stressors they have experienced in a specified period of time. The concern is that these measures contain items that are confounded with the psychological adjustment measures they are predicting (L. H. Cohen et al., 1987; B. S. Dohrenwend & Dohrenwend, 1974; B. S. Dohrenwend et al., 1984). Specifically, life stressors measures often include items about negative experiences over which individuals have some control. To the extent that life stressors items reflect preexisting levels of psychological adjustment, they represent events that are *controllable* or *dependent* stressors, events, and circumstances that are at least partially dependent on the behavior of the individual (L. H. Cohen et al., 1987; Masten et al., 1994). As a result, some suggest that the life stressors–psychopathology associations reported in much of the research are inflated.

Researchers have dealt with this measurement problem in a variety of ways. One approach is to study only uncontrollable, or fateful, stressors, those that occur independently of an individual's actions (Shrout et al., 1989). Findings using this method have been inconsistent, with some studies finding an association with psychological adjustment (DuBois, Felner, Brand, Adan, & Evans, 1992; Newcomb, Huba, & Bentler, 1986) and others not (Dubow, Tisak, Causey, Hryshko, & Reid, 1991; Gersten, Langner, Eisenberg, & Simcha-Fagan, 1977). A second approach examines the role of controllable and uncontrollable life stressors simultaneously, but here again the findings have been inconsistent. Cohen and his colleagues (Burt, Cohen, & Bjorck, 1988; L. H. Cohen et al., 1987) found that only controllable major stressors were related to subsequent psychological symptoms, whereas others have found this association for both controllable and uncontrollable stressors (Masten et al., 1994; Rowlison & Felner, 1988). Thus, the extent to which controllable and uncontrollable stressors are related to depressive symptoms remains unclear.

It should be noted that relying on research participants' reports of life stressors has, in and of itself, been called into question. In fact, any stressors self-report could be considered dependent in that it reflects what the individual perceives to be significant experiences (B. S. Dohrenwend et al., 1984; Rudolph et al., 2000). We argue that additional research using this type of measure is important for two reasons. First, the vast majority of previous studies, as well as many recent studies, have used this approach in spite of the methodological critiques (Cole, Nolen-Hoeksema, Girgus, & Paul, 2006; Kim, Conger, Elder, & Lorenz, 2003; Morales & Guerra, 2006). Second, the primary goal of this study is to contribute to a better understanding of the S-DS process, which is at least partially based on research using this method of life stressors measurement.

Alternatives to Social Causation Models

Important questions have been raised regarding the strength and direction of the life S-DS relationship suggested by social causation models. At least three alternate theory-based psychosocial models could represent the processes at play.

Although not generally acknowledged in the S-DS literature, an alternate unidirectional model can be derived from the mood congruent memory research. This research has established that an individual's mood is likely to influence the types of memories recalled. This *mood-congruent cognitive bias* model suggests that the association between depressive symptoms and reports of stressful life circumstances is due, at least in part, to the fact that people are more likely to recall negative memories when in a negative mood and positive memories when in a positive mood (for a review, see Blaney, 1986; see Figure 2a).

Mood congruence studies have used both observational and experimental designs. One type of observational design involves assessing participants' existing mood, having them learn novel material, and then comparing more and less depressed participants on the emotional valence of material recalled. These studies find that more depressed participants recall more negatively valenced material than nondepressed participants (Ingram, Smith, & Brehm, 1983; McDowall, 1984). Experimental studies in which researchers induce a depressed mood in the treatment group have also found evidence of mood-congruent cognitive bias in recall (Bower, 1981; Klaassen, Riedel, Deutz, & van Praag, 2002).

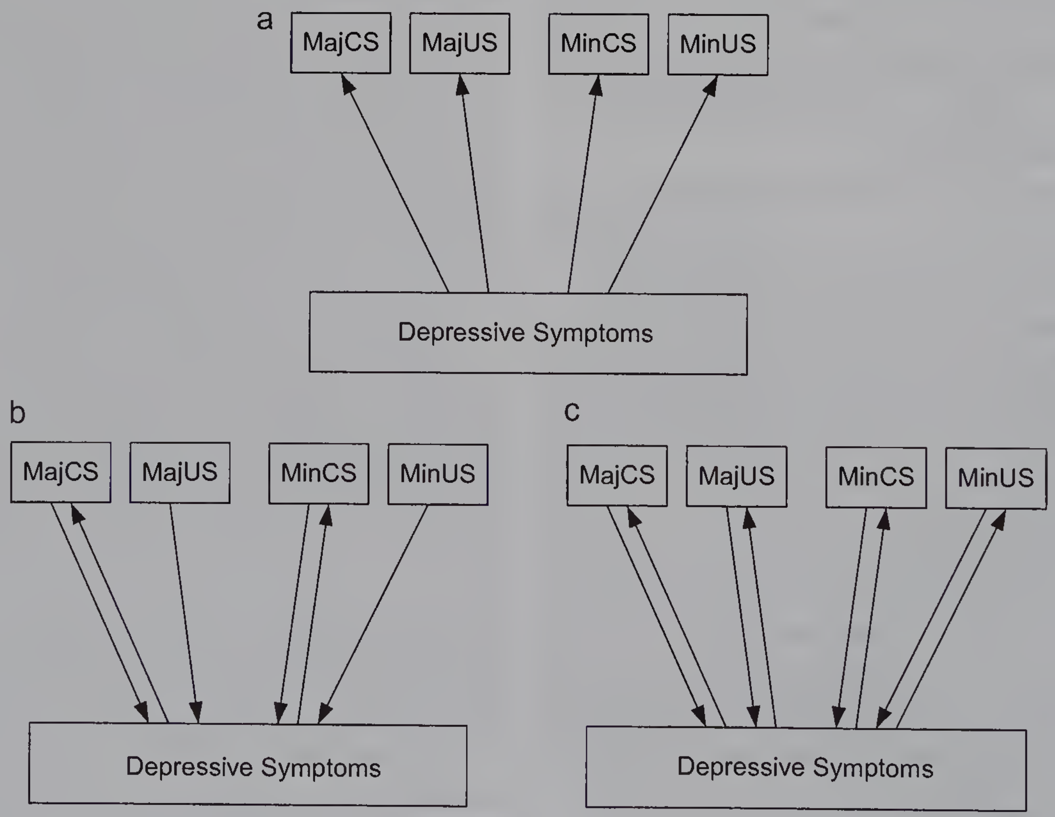


Figure 2. Heuristic models of the alternate theories.
Note. MajUS = major uncontrollable stressor; MinUS = minor uncontrollable stressor; MajCS = major controllable stressor; MinCS = minor controllable stressor.

Other researchers argue that unidirectional models do not account for the transactional relationships between individuals and their environments (Cicchetti, 1984; Felner & Felner, 1989; Sameroff & Mackenzie, 2003). An example of this type of transactional process is the following: The death of a parent causes an adolescent to become depressed; the resulting emotions and behaviors lead to an increase in stressors, such as interpersonal problems with friends, and these stressors cause the adolescent to become more depressed. Early research investigating transactional S-DS relationships compared the regression coefficient for a stressors-to-depressive symptoms effect (S → DS) estimated in one regression to the coefficient of a depressive symptoms-to-stressors effect (DS → S) estimated in a separate analyses, with some studies finding that the S → DS associations was stronger (DuBois et al., 1992; Pianta & Egeland, 1994) and others the reverse (L. H. Cohen et al., 1987).

Hammen and her colleagues are among those who propose a transactional S-DS model (Hammen, 1991, 1992a; Rudolph et al., 2000). While recognizing that depression may be triggered by stressful circumstances that a person cannot control, Hammen's (1991) *stress generation model* suggests that depressed people behave in ways that create stressful circumstances in their lives (see Figure 2b). Using an in-depth interview method designed to control for the influence of cognitive bias on event recall, she and her colleagues have found that depressed women and adolescents recall more controllable stressors, but not more uncontrollable stressors, than their nondepressed counterparts (Hammen, 1991; Rudolph et al., 2000).

Studies using self-report checklist measures of stressors have also found support for the stress generation theory (Cole et al., 2006; Kim et al., 2003). However, these studies did not account for the possibility that the stressors measures might include both controllable and uncontrollable stressors or the possibility that the DS → S associations might represent mood-congruent cognitive bias. In fact, in order to determine whether stress generation and/or cognitive bias account for significant DS → S paths, research must distinguish between controllable and uncontrollable stressors. Specifically, the presence of significant DS → S paths only for controllable stressors would support the conclusion that stress generation is at work. On the other hand, significant DS → S paths for uncontrollable stressors would indicate that mood-congruent cognitive bias is at work, as these stressors are not likely to result from stress generation. It should be noted that if mood-congruent cognitive bias is operating, it is likely to produce significant DS → S associations for both controllable and uncontrollable stressors, as its influence would apply to recall of both types of stressors. For research using self-report measures of stressors, determining what portion of the depressive symptoms-to-controllable stressors association is due to cognitive bias and what portion is due to stress generation would not be possible.

A second possible transactional model is a *fully transactional* model that simultaneously accounts for the processes suggested by all models discussed thus far: the contribution of life stressors to the development of depressive symptoms, the influence of current mood on one's recall of negative experiences, and the possibility that depressed individuals behave in ways that create stressful life circumstances (see Figure 2c). A formal test of whether the fully transactional model provides the best fit would contribute to our understanding of how life stressors influence the development of depressive symptoms in at least two important ways. First, it would demonstrate whether stress generation is the only transactional process operating or whether mood-congruent recall bias should be considered. Second, it would provide a rigorous test of the extent to which stressful life circumstances predict the development of depressive symptoms because any effects would demonstrate the S → DS effect net of stress generation and/or cognitive bias.

Current Study

The primary goal of the present study was to systematically estimate and compare the fit of six competing theory-based psychosocial models of the S-DS relationship that have been proposed in previous research. Furthermore, to address a central methodological critique of existing life stress research, we used measures of major and minor life stressors that distinguish between controllable and uncontrollable experiences in the estimation of the models.

Method

Sample

Data for this study come from the Adolescent Pathways Project (APP), a longitudinal study of youth attending public schools in Baltimore, MD; Washington, DC; and New York, NY (Seidman, 1991). The APP study was designed to apply an ecological framework to the study of adolescent development among ethnically diverse youth attending schools with high proportions of students from low-income families. In the 1st year of the study, we assessed participants' ($N = 907$) race/ethnic background by asking them to indicate whether they were White/not Hispanic (19%), Black/not Hispanic (32%), Hispanic (40%), Asian (5%), or more than one race/ethnic group (4%). In the 2nd year of the study, we asked participants to describe in their own words their families' ethnic background(s) and found that most Hispanic students' families were originally from Puerto Rico or the Dominican Republic, most Black students' families were originally from the Caribbean islands or were African American, and most White students' families were originally from Italy or Greece. Fifty-four percent of the participants were female and the average age at Wave 1 was 11.40 years ($SD = 0.92$), with the large standard deviation representing the fact that the sample spanned two grades. An examination of 1990 U.S. Census data for the 707 New York participants indicated that the majority lived in census tracts where at least 20% of the households had incomes below the federal poverty threshold. Schools with a minimum of 60%–80% of the student body eligible for reduced price or free lunch were targeted for participant recruitment. Within these schools, all students in the grade prior to the transition to a middle grades school were contacted for participation in the study (i.e., the fifth or sixth grade, depending on school grade structure).

Active parental consent was required for participation. APP staff provided a study description and consent form (in both English and Spanish) for teachers to send home with children, which children were to give to parents and return to school. A follow-up notice was provided to teachers approximately 1 week later. The 907 early adolescents for whom valid data were collected at Wave 1 represented 38% of the adolescents recruited for participation. We believe it is likely that the numerous steps and individuals involved in the active consent procedure contributed to this participation rate. Sixty-four ($N = 581$) percent of the original sample participated in all three waves of data collection and 22% participated in two waves of data collection (Wave 1 and either Wave 2 or 3, $N = 183$).

Attrition analyses comparing adolescents who participated at all three waves to those who are missing one or more waves of data indicate that the two groups differed in terms of race/ethnic group composition and age, but that the effect sizes were small. A comparison of the three largest race/ethnic groups showed that a larger percentage of the White adolescents (74%) participated at all three waves than Latino (66%) or Black (56%) adolescents, $\chi^2(2) = 16.79$, $p < .001$. However, a binary logistic regression in which retention at all three waves (vs. missing any wave) was regressed on two variables indicating whether a participant was Black or White (with the largest group, Hispanic, as the reference group) indicated that only 2% of the variance in attrition status is associated with an adolescent's race/ethnicity. In terms of age, adolescents missing one or more waves of data were slightly older than those with data for three waves ($M = 11.49$ years, $SD = 1.00$ and $M = 11.35$, $SD = 0.86$, respectively), $t(902) = 2.18$, $p < .01$. The effect size for this age difference was small, Cohen's $d = .15$ (J. Cohen, 1988). There was not a significant difference in the gender composition of the two groups, $\chi^2(1) = 3.32$, *ns*. We present the attrition analyses for study variables after describing the measures.

The APP collected Wave 1 data in the spring of 1988, which was participants' last year of elementary school. Waves 2 and 3 data collection took place approximately 1 and 2 years later, respectively. Small teams of APP staff from diverse racial and ethnic backgrounds collected data in group settings at schools. Instructions and questions were read aloud to the group while staff members circulated throughout the room to answer participants' questions and spot-check responses to measures with more complicated directions.

Measures

Major stressors. Respondents reported negative major stressors using a checklist measure consisting of 17 items adapted from Johnson and McCutcheon's (1980) Life Events Checklist and 9 additional items representing stressors that that inner-city adolescents might experience. Participants indicated whether they had experienced each event in the past year and scores represent the number of stressors adolescents reported. We describe the procedure for distinguishing controllable and uncontrollable items below. Because stressors scales are designed to assess experiences that may occur independently from each other, it is not appropriate to estimate the internal consistency of these measures (Hoffman & Su, 1997; Kim et al., 2003).

Minor stressors. Respondents reported minor stressors using the 27-item Daily Hassles Scale (Seidman et al., 1995), which consists of items adapted from Rowlison and Felner's (1988) daily hassles measure and additional items developed to assess resource-related and neighborhood stressors. Respondents rated each stressor on a 5-point scale, indicating either that the stressor, or hassle, *had not occurred in the past* *year* (1), *or that it had occurred and was either not a* (2), *a little bit* (3), *a moderate* (4), or *a very big* (5) hassle. We dichotomized each item score as either a nonoccurrence (scores of 1 or 2) or an occurrence (scores of 3, 4, or 5) and computed a composite score of the number of minor stressors experienced in the past month. The procedure used to identify controllable and uncontrollable minor stressors is described below. These scales are designed to assess experiences that may occur independently from each other and it is not appropriate to estimate their internal consistency.

Depressive symptoms. The depressive symptoms scale consists of 14 items from Achenbach and Edelbrock's (1987) Youth Self-Report. Respondents reported whether each item was *not true* (0), *somewhat or sometimes true* (1), or *very true or often true* (2) for them in the past 6 months. Internal consistency was adequate (αs = .84, .84, and .86 in Waves 1, 2, and 3, respectively).

Identifying Controllable and Uncontrollable Stressors

Because neither stressors measure was designed to distinguish controllable and uncontrollable stressors, we conducted a preliminary study to inform this distinction. We based our methodology on that used by B. S. Dohrenwend et al (1984) in their research investigating conceptual and operational distinctions for measurements of psychological symptoms, major stressors, and minor stressors.

A convenience sample of psychologists and advanced psychology graduate students (N = 27, 63% White, 19% Asian, 11% Black, and 7% Hispanic) rated the extent to which adolescents may play a role in experiencing each of the negative major and minor stressor items. Interspersed with the stressor items were 10 anchor items designed to clearly represent either a controllable event (e.g., *watched TV, read a magazine*) or an uncontrollable event (e.g., *it snowed*). Respondents rated items using a 5-point scale, ranging from 1 (*an adolescent almost certainly plays some role in this event*) through 5 (*an adolescent almost certainly plays no role in this event*), with 3 (*an adolescent may or may not play some role in this event*) as the midpoint.

We conducted one-sample t tests, using the midpoint (3.0) as the criterion, to determine whether an item was controllable or uncontrollable. We classified items with means significantly greater than 3.0 as uncontrollable stressors and those with means significantly less then 3.0 as controllable. This process identified 6 controllable major stressor items and 17 uncontrollable items, and eliminated the 3 major stressor items with means not significantly different from 3.0. For minor stressors, we classified 9 items as controllable, 11 as uncontrollable, and eliminated 7 with means not significantly different from 3.0. (A full list of all items, their means and standard errors, and classifications is available from the authors.) Means for the items designed to anchor ends of the scale were significantly different from 3.0 in the directions anticipated, with means for the controllable items below 1.5 and means for the uncontrollable items equal to 5.0.

Descriptive analyses for the new controllable and uncontrollable subscale scores showed a low endorsement rate for the controllable major stressors measures at all three waves (e.g., became pregnant, expelled from school). This construct was excluded from all subsequent analyses. Means and standard deviations of all study variables are reported in Table 1.

Additional Attrition Analyses Examining Study Variables

We conducted additional attrition analyses on Wave 1 scores of all study variables comparing adolescents who participated in all waves with those who did not. In order to adjust for the two demographic characteristics significantly associated with attrition (identified above), we conducted a 2 × 3 (Attrition Group × Race/Ethnic Group) analysis of covariance with age entered as a covariate. There were significant differences between the two attrition groups on two of the four study variables. The depressive symptoms score was lower for the group that participated in all three waves than for the group missing one or two waves of data ($M_{\text{adjusted}} = 0.47$, $SE = .02$ compared to $M_{\text{adjusted}} = .52$, $SE = .02$), $F(1) = .4.52$, $p < .05$, $\eta^2 = .01$. There was also a significant difference between the two groups for major uncontrollable stressors, with the group participating in all three waves reporting fewer stressors than the group that did not ($M_{\text{adjusted}} = 1.25$, $SE = .06$ compared to $M_{\text{adjusted}} = 1.48$, $SE = .09$), $F(1) = 4.06$, $p < .05$, $\eta^2 = .005$). In both cases, the percent of variance explained was less than 1%.

Results

All structural equation models (SEMs) were estimated using the full information maximum likelihood (FIML) estimation method available in Mplus, version 4.1 (Muthén & Muthén, 2006). We chose FIML because this method of estimation uses all information in the observed data, including cases with missing data. This approach to handling missing data is preferable to listwise deletion of cases, which

Table 1
Pairwise Correlations, Means, and Standard Deviations of Model Variables

Variable	1	2	3	4	5	6	7	8	9	10	11	12
1. MajUS1	1.00											
2. MinCS1	.14**	1.00										
3. MinUS1	.08*	.54**	1.00									
4. DS1	.18**	.45**	.42**	1.00								
5. MajUS2	.21**	.14*	.12**	.17**	1.00							
6. MinCS2	.10**	.39**	.32**	.34**	.19**	1.00						
7. MinUS2	.05	.32**	.48**	.35**	.16**	.52**	1.00					
8. DS2	.08	.27*	.33**	.57**	.13**	.46**	.47**	1.00				
9. MajUS3	.21**	.07	.09*	.09**	.31**	.19**	.21**	.15**	1.00			
10. MinCS3	.08**	.34**	.23**	.28**	.15**	.47**	.33**	.36**	.22**	1.00		
11. MinUS3	.09*	.25**	.36**	.28**	.13**	.34**	.51**	.37**	.19**	.55**	1.00	
12. DS3	.13**	.22**	.24**	.47**	.11**	.29**	.31**	.63**	.15**	.46**	.43**	1.00
M	1.33	3.49	3.23	0.51	1.46	3.25	2.67	0.40	1.42	3.21	2.62	0.36
SD	1.32	2.16	2.35	0.35	1.52	2.14	2.29	0.33	1.45	2.23	2.30	0.33

Note. MajUS = major uncontrollable stressor; MinCS = minor controllable stressor; MinUS = minor uncontrollable stressor; DS = depressive symptoms; 1 = Wave 1; 2 = Wave 2; 3 = Wave 3.
$^{*}p < .05$. $^{**}p < .01$.

results in both the loss of potentially valuable information and reduced statistical power (Schafer & Graham, 2002). We conducted all analyses a second time after imputing missing values using the expectation maximization algorithm and found that the results were essentially the same for the two methods of handling missing data. We report the results using FIML estimation below.

Given this study's large sample size, the chi-square statistic is not diagnostic of model fit and we used the two-index strategy recommended by Hu and Bentler (1998, 1999) instead. The two fit indices we used were the standardized root mean square residual (SRMR), which Hu and Bentler recommend for all SEM analyses, and the root mean square error of approximation (RMSEA) because it is sensitive to complex model misspecification (Hu & Bentler, 1998). Both indices estimate the residual between the estimated covariance matrix and the sample covariance matrix. Hu and Bentler (1998) suggest .08 as the cutoff value for SRMR and .06 for RMSEA. We used the chi-square difference statistic (χ^2_{diff})—the difference between the chi-square of two models—to compare the fit of nested models and the Akaike information criterion (AIC) to compare the fit of nonnested models (Kline, 2005).

Each of the models contains a set of identical paths that estimate the autoregressive associations and certain residual covariance estimates. The autoregressive paths represent the stability of each variable across the three data points. We allowed the residuals for the controllable and uncontrollable minor stressors scores to covary within each wave because they were originally from the same instrument. The means, standard deviations, and intercorrelations of all model variables appear in Table 1. (Given our use of FIML estimation, these correlations are not technically those of the data used in the SEM analyses.)

Social Causation Models

Independent effects model. This model included direct S → DS paths for all stressors at each wave. The model does not provide an adequate fit of the data, SRMR = .099, RMSEA = .072. Table 2 presents the fit statistics for this and all subsequent models, and Table 3 presents the estimated coefficients for all models.

Although it is not generally meaningful to interpret a model that does not adequately fit the data, we will briefly review this model because it is most like the analysis conducted in much of the S-DS research. It is important to note, however, that even with this model's similarities to analyses in previous studies, it is considerably more conservative because it predicts change in depressive symptoms using change in stressors not the overall level of stressors (i.e., a single wave of stressors data).

The autoregressive coefficients in this model demonstrate that there is considerable stability in stressors and depressive symptoms over the course of the study, with most of the βs representing moderate to strong associations. Regarding the S → DS paths, all three stressors are significantly related to depressive

Table 2

Summary of Fit Statistics for Hypothesized Models of Life Stressors of Depressive Symptoms

Models	SRMR	RMSEA	AIC	χ^2	*df*	χ^2_{diff}	df_{diff}
Social causation models							
1. Independent effects model	.099	.072	26,717.29	248.31	44		
2. Mediated model	.086	.067	26,691.22	220.24	43		
Δ Model 1 to Model 2						28.07***	1
3. Partially mediated model	.085	.068	26,684.01	207.04	40		
Δ Model 2 to Model 3						13.20***	3
Alternatives to social causation models							
4. Cognitive bias model	.056	.050	26,614.13	144.15	44		
Δ Model 1 to Model 4							
5. Stress generation model	.084	.065	26,681.81	204.83	42		
6. Fully transactional model	.052	.051	26,569.18	129.21	38		
Δ Model 4 to Model 6						14.94*	6

Note. $N = 907$. SRMR = standardized root mean square residual; RMSEA = root mean square error of approximation; AIC = Akake information criterion.
$^*p < .05$. $^{***}p < .001$.

symptoms at Wave 1, a finding that is in line with existing cross-sectional studies. The coefficients for the S→DS paths in Waves 2 and 3 suggest that only changes in the minor stressors are positively associated with changes in depressive symptoms. Although the nonsignificant major stressors – depressive symptoms paths appear to contradict much of the published research regarding major life stressors, this may be due to the fact that the change is being modeled at Waves 2 and 3 not the statistical relationship with a single cumulative measure.

For the independent effects model and all subsequent models, we anticipated that the S-DS association would be strongest between depressive symptoms and the most recently experienced stressors. By estimating the association between stressors and depressive symptoms reported at the same data collection point, we were in effect examining the association between stressors experienced in the preceding year and recent depressive symptoms. However, to investigate whether the models should include lagged paths between stressors experienced in the prior year and recent depressive symptoms, we estimated two alternate versions of the model. The first model estimated the associations between previously experienced stressors (e.g., Wave 1) only and change in depressive symptoms (e.g., Wave 2). This fit of this model was worse than that of model estimating the S-DS association for recently occurring stressors. The second model added the lagged S→DS paths to the original independent effects model, so that the model included both the lagged and the concurrent paths. The fit of this model was also inferior to the originally estimated model. Based on these two analyses, we determined that modeling paths between the measures of most recent stressors and depressive symptoms is the best approach for testing the social causation theories. (Model fit statistics and estimated coefficients for the alternate models may be obtained from the authors.)

Fully mediated model. For this model, we removed the paths from uncontrollable major stressors to depressive symptoms at each wave and replaced them with mediated pathways through each of the minor stressors. In spite of the multiple structural differences in this model compared to the independent effects model, the degrees of freedom have decreased only by one (44 – 43). This is because the model simultaneously gained three degrees of freedom from removing the major stressors to depressive symptoms paths at each wave, but lost four through the addition of paths from uncontrollable major stressors to the two minor stressors measures at Waves 2 and 3. It should be noted that because SEMs estimate the noncausal associations (correlations) among all purely exogenous variables, changing the paths at Wave 1 from noncausal to causal did not effect the degrees of freedom for this part of the model. The fit of the mediated model is significantly better than the independent effects model, $\chi^2_{\text{diff}}(1) = 28.07, p < .001$, but does not provide an adequate fit to the data, SRMR = .086, and RMSEA = .067.

Partially mediated model. For the partially mediated model, we added direct paths from uncontrollable major stressors to depressive symptoms at each wave. The model does not provide an adequate fit of the data, SRMR = .085 and RMSEA = .068. However, the fit is significantly better than the fully mediated

Table 3
Estimated Standardized Coefficients for All Models

	Social causation models				Transactional models	
Coefficient	Ind. effects	Med. effects	Part. med. effects	Cog. bias	Stress gen.	Fully transact'l
Wave 1						
MajUS→MinCS		.16**	.15***			
MajUS→MinUS		.09**	.09*			
MajUS→DS	.11***		.11**		.11***	
MinCS→DS	.30***	.31***	.30***		.30***	
MinUS→DS	.26***	.26***	.26***		.26***	
DS→MajUS				.18***		.18***
DS→MinCS				.45***		.45***
DS→MinUS				.42***		.42***
Wave 2						
MajUS→MinCS		.15***	.15***			
MajUS→MinUS		.12***	.12***			
MajUS→DS	−.01		−.01		.01	-.15**
MinCS→DS	.20***	.19***	.20***		-.08	.01
MinUS→DS	.23***	.23***	.23***		.36***	.15**
DS→MajUS				.12**		.26***
DS→MinCS				.38***	.32***	.33***
DS→MinUS				.37***		.25***
Wave 3						
MajUS→MinCS		.16***	.16***			
MajUS→MinUS		.10**	.10**			
MajUS→DS	.01		.01		.03	−.06
MinCS→DS	.22***	.23***	.22***		.01	.08
MinUS→DS	.13***	.13***	.13***		.23***	.03
DS→MajUS				.12**		.17**
DS→MinCS				.36***	.24***	.30***
DS→MinUS				.30***		.26***
Wave 1 → Wave 2 autoregressive paths						
DS→DS	.44***	.44***	.44***	.57***	.48***	.54***
MajUS→MajUS	.23***	.23***	.23***	.22***	.23***	.21***
MinCS→MinCS	.36***	.35***	.35***	.28***	.29***	.29***
MinUS→MinUS	.44***	.44***	.44***	.37***	.47***	.39***
Wave 2 → Wave 3 autoregressive paths						
DS→DS	.50***	.50***	.50***	.62***	.53***	.59***
MajUS→MajUS	.32***	.33***	.33***	.31***	.32***	.30***
MinCS→MinCS	.44***	.42***	.42***	.36***	.38***	.37***
MinUS→MinUS	.47***	.46***	.46***	.41***	.49***	.42***
R^2						
MajUS1				.03		
MinCS1		.02	.02	.20		
MinUS1		.01	.01	.18		
DS1	.26	.25	.26		.26	.26
MajUS2	.05	.05	.05	.07	.05	.04
MinCS2	.13	.15	.15	.27	.25	.27
MinUS2	.20	.21	.21	.32	.22	.32
DS2	.39	.39	.39	.32	.35	.36
MajUS3	.10	.11	.11	.12	.11	.11
MinCS3	.19	.20	.20	.33	.29	.33
MinUS3	.22	.23	.23	.33	.24	.33
DS3	.41	.41	.41	.38	.40	.41

Note. MajUS = major uncontrollable stressors; MinCS = minor controllable stressors; MinUS = minor uncontrollable stressors; DS = depressive symptoms; Ind. Effects = independent effects model; Med. Effects = mediated effects model; Part. Med. Effects = partially mediated effects Model; Cog. Bias = mood-congruent cognitive bias model; Stress Gen. = stress generation model; Fully Transact'l = fully transactional model. Unstandardized coefficients and their standard errors may be requested from the authors.
*$p < .05$. **$p < .01$. ***$p < .001$.

model, $\chi^2_{\text{diff}}(3) = 13.20, p < .001$, in spite of the fact that all the direct paths from uncontrollable major stressors are not significantly different from zero. In fact, a closer examination of the mediated models suggests that mediation is not occurring even though both models provide a better fit to the data than the independent effects model. First, even in the independent effects model, the Waves 2 and 3 direct paths from uncontrollable major stressors to depressive symptoms were not statistically significant. Second, the coefficients from the minor stressors to depressive symptoms remain essentially the same across the three social causation models, indicating that the strength of their associations with DS does not increase over the unique contributions estimated in the independent effects model.

Mood-Congruent Cognitive Bias Model

The structure of the cognitive bias model is essentially the opposite of the independent effects social causation model: all S→DS paths were removed and replaced with DS→S paths. To estimate a model that would be similar to the independent effects model except for the direction of the S-DS paths, we estimated the covariance between the residuals of Wave 1 stressors. This was necessary because as exogenous variables in the independent effects models they were automatically estimate.

In contrast to the social causation models, this model provides an adequate fit to the data, SRMR = .056 and RMSEA = .040. Because this model is not nested with the social causation models, we used the AIC to compare model fit. The lower AIC statistic for the mood-congruent cognitive bias model indicates that it provides a better fit to the data than any of the social causation models, for example, $\text{AIC}_{\text{cognitive bias}} = 26{,}313.13$ and $\text{AIC}_{\text{independent effects}} = 26{,}717.29$. All the coefficients for the DS→S paths are statistically significant, with standardized coefficients ranging from 12 to .18 for major stressors and .30 to .45 for minor stressors. While changing the direction of structural paths in an SEM can, under particular circumstances, produce an equivalent model, this is not the case when changing the direction of the S-DS paths to specify the mood-congruent cognitive bias model. Equivalent models are those that produce the exact same estimated covariance matrix and fit statistics even though some of the structural paths in the models are different. For a structural change in a model to produce an equivalent model, the predictors of the initially endogenous variable (e.g., DS in the independent effects model) must include the predictors of the exogenous variable in the original model (e.g., S in the independent effects model). Only under these circumstances will the respecified model produce an identical estimate of the covariance matrix (Kline, 2005; Lee & Hershberger, 1990).

As we did in the case of the independent effects model, we conducted an additional set of analyses in order to rule out the possibility that including paths representing lagged associations between depressive symptoms and stressors would result in a better fitting model. We followed the same procedure to test the alternate models that we used for the independent effects model: The first model estimated the associations between previously reported depressive symptoms (e.g., Wave 1) and change in stressors (e.g., Wave 2), and the second model added the lagged paths to the original model. The fit of both models was worse than that of the original model. (Model fit statistics and estimated coefficients for the alternate models may be obtained from the authors.)

Stress Generation Model

The stress generation model includes S→DS paths from all stressors and DS→S paths only for controllable minor stressors (to represent the generation of controllable stressors). Although the DS→S paths representing stress generation are statistically significant, this model does not provide an adequate fit to the data, SRMR = .084 and RMSEA = .065. Given that all the DS→S paths were significant in the mood-congruent cognitive bias model, it is not surprising that the stress generation model, which excludes four of those paths, does not fit as well.

Fully Transactional Model

The fully transactional model integrates components of the mood-congruent cognitive bias, stress generation, and independent effects social causation models. (We did not include the mediation paths because the results did not demonstrate mediation.) At Waves 2 and 3, the fully transactional model includes S→DS paths from all stressors and DS→S paths to all stressors. At Wave 1, we included the unidirectional paths representing the mood-congruent cognitive bias model. The model provides an adequate fit of the data, SRMR = .052 and RMSEA = .051, and the fit of this model is significantly better than that of the mood-congruent cognitive bias model, $\chi^2_{\text{diff}}\ (6) = 14.95, p < .05$, indicating that it provides the best fit of all models examined.

As was the case with the mood-congruent cognitive bias model, all the DS→S paths are significant, with most having a moderate effect size. Only two of the S → DS paths are statistically significant, both in Wave 2, which is the 1st year in a junior high/middle school for all study participants. The significant positive S → DS path for minor uncontrollable stressors indicates that experiencing an increase in minor uncontrollable hassles predicts an increase in depressive symptoms for these adolescents, net of the DS → S effect, $\beta = .15$. The standardized path coefficient for the S → DS path for uncontrollable major stressors at Wave 2 has changed from essentially nil (−.01) in the social causation models to −.15.

Discussion

The primary goals of this study were to compare the fit of six theory-based, psychosocial models of the stress–depressive symptoms association and to address a central methodological critique of this line of research—that self-report life stressors measures may be confounded with psychological symptoms. Our SEM analyses demonstrated that the social causation models, which represent the most common theoretical approach to studying the S-DS association, do not provide an adequate fit to the data in spite of the fact that most of the S → DS paths were statistically significant. Had we used an alternate data analytic method, such as ordinary least squares regression, and examined only this theoretical model, we could have incorrectly concluded that the process accounting for the S-DS associations is social causation alone. Only when we modeled the influence of mood on recall—either alone in the mood-congruent cognitive bias model or in the fully transactional model—did the estimated models fit the data adequately.

We will discuss the findings from the social causation models in order to compare our findings to those from previous research. The findings in the cross-sectional (Wave 1) portion of these models were similar to those of previous research in that the S → DS paths were significant for major and minor stressors. Comparing our longitudinal findings to previous research is less straightforward because the structure of the longitudinal components of our models tested the extent to which a *change* in stressors predicts a *change* in depressive symptoms, although most previous longitudinal S-DS studies have predicted depressive symptoms using only a single, cumulative measure of stressors: either a prior report of stressors (Hammen, Burge, & Adrian, 1991; Morales & Guerra, 2006; for a review, see Grant, Compas, Thurm, McMahon, & Gipson, 2004) or a recent report of stressors (Compas, Slavin, Wagner, & Vannatta, 1986; DuBois et al., 1992, 1994; Nolen-Hoeksema & Girgus, 1994). Research using either design has consistently found that major and minor stressors predict increases in depressive symptoms. Our approach was considerably more stringent in that we examined whether *changes* in stressors predict changes in depressive symptoms.

Given that many previous studies have found either a direct or mediated relationship between major stressors and depressive symptoms, the fact that our models did not was unexpected. There are at least three explanations for this finding. One possibility is that this more stringent test demonstrates that a single, cumulative measure of major stressors predicts change in depressive symptoms, but an increase in major stressors does not. A second possibility is that the S-DS associations found in other studies were due to the fact that the stressors measures included items confounded with psychological adjustment. A third possibility is that the magnitude of the major stressors → DS association is smaller for low-income, urban adolescents than for the adolescent participants in previous studies.

The most important finding from the social causation models is that they do not provide an adequate fit of the data even though almost all estimated parameters were statistically significant. This finding demonstrates that examining only social causation models of the S-DS process is not appropriate when using self-report measures, and we suggest that any S-DS study incorporating a social causation perspective should go on to rule out other possible explanations for the S-DS association. Significant S → DS paths are not, in and of themselves, an adequate stopping place for these investigations. The fact that the mood-congruent cognitive bias model provided a markedly better fit of the data indicates that S-DS research must take into account the likelihood that life stressors reports reflect participants' mood.

In light of the fact that several recent studies have found support for stress-generation models of the S-DS association (Cole et al., 2006; Hammen, 1991; Kim et al., 2003; Rudolph et al., 2000), one might question our finding that the stress generation model did not fit the data. We suggest that important methodological differences in our model account for the divergent results, and that our findings make important theoretical contributions to understanding the S-DS process. As we discussed earlier, without distinguishing between and including both controllable and uncontrollable stressors, it is not possible to conclude that stress generation per se is responsible for significant DS → S paths. In order to conclude that only stress generation is at play requires demonstrating that depressive symptoms *do* predict increases in controllable stressors and *do not* predict increases in uncontrollable stressors. This was our approach. The stress generation model included DS→S paths for controllable stressors only and did not provide an adequate fit to the data, these paths were significant. The fully transactional model included DS→S paths for all stressors, found that all of these paths were significant, and provided an adequate fit to the data.

Although the stress generation model we estimated did not fit the data, we do not rule out the validity of this model. As discussed earlier, there are alternate approaches to measuring life stressors that distinguish between controllable and uncontrollable stressors, with the work Hammen, Rudolph, and their colleagues exemplifying this approach (Hammen, 1991; Rudolph et al., 2000). In these studies, researchers conducted interviews designed to elicit details about the context in which stressful experiences occurred, and then assessed the degree to which a typical person would be effected by such an experience (how stressful the event was) and the degree to which the participant contributed to the event's occurrence (how controllable the event was). Their assumption is that these stressors measures do not include mood-congruent cognitive bias and, as a result, do not need to statistically model this process. Using this method, Hammen and her colleagues have consistently found that depressed individuals are more likely to experience interpersonal, dependent, major stressors (Hammen, 1991; Rudolph et al., 2000).

A second reason not to rule out the validity of the stress generation model is that the self-report checklist stressors measures we used did not allow us to determine the extent to which the DS→controllable stressors paths represent mood-congruent cognitive bias and to what extent they represent stress generation. It is possible that these paths represent both processes.

Although the stress generation model did not provide an adequate fit to the data in this study, our findings regarding the S-DS process are similar to those of Hammen and her colleagues in that we demonstrate one way in which cognition functions in the stress-depression process (Hammen, 1992b; Rudolph, 2002; Rudolph & Hammen, 1999; Rudolph et al., 2000). They propose that children who encounter a great deal of uncontrollable stress early in their lives develop the belief that they cannot exert control over their environments, which results in the development of negative cognitive styles. Manifestations of this are the negative interpretation of events and the tendency to be more likely to recall the negative experiences in one's life. It is important to note that depressed individuals are not constructing negative memories of events that never happened, but that they are more likely to both remember these experiences and interpret them more negatively.

Before we discuss the results of the fully transactional model, we will address the composition of the sample in this study, as this informs our discussion of the model. As is often the case in development studies, adolescents who participated in this study are not representative of the population of adolescents. In fact, most previous studies of the S-DS process have focused on the experiences of White adolescents (Clements, 2003). The overarching goal of the sampling strategy used by the APP study, from which data for the current study were drawn, was to address the fact that developmental scientists have much to learn about the factors that promote and hinder positive academic and social-emotional development among low-income urban adolescents. Furthermore, the group that participated in the study represents the subset of this population with characteristics (whatever they may be) that enabled them to obtain their parents' consent and to participate.

Although we recognize that the sample in this study represents a selected group of adolescents, we anticipate that the mood-congruent cognitive bias effect is generalizable to adolescents from other socioeconomic and demographic backgrounds because previous experimental and quasi-experimental studies have demonstrated that it is a robust phenomenon. Although, as for any finding, replication in studies of other adolescent populations is necessary.

We are more hesitant to draw strong conclusions about the S→DS effects. The nonsignificance and small effect size of many of these coefficients may be due to a number of factors. One issue, as discussed earlier, is that the sample is not representative. It is possible that the findings of the fully transactional model are an accurate representation of the S→DS process for this particular population of adolescents. However, without other studies, using this design we cannot know whether the S→DS processes identified here are unique to this group of adolescents. It could also be that early adolescence is not the developmental period during which life stressors are most likely to cause depressive symptoms. For example, it may be that the relationship between stressful life experiences and depressive symptoms is stronger at earlier developmental periods, such as childhood, and that persistent symptoms cause adolescents to perceive their experiences more negatively. It is also possible that the S→DS effects cannot be detected due to measurement error. For example, Shrout et al. (1989) point out that the impact of any stressor reported on checklist measures may vary considerably for an individual depending on context and circumstances in which the stressor occurs. Furthermore, it is important to note that the fully transactional model provided a significantly better fit to the data than the cognitive bias model, suggesting that the S→DS paths are a meaningful component of the model and processes that research should continue to explore.

An additional consideration is that the S→DS path coefficients are smaller than the DS→S coefficients due to the structure of the analyses. Because multiple stressors measures are used to predict depressive symptoms, the paths reflect each stressor's unique association with symptoms. In contrast, the paths representing the effect of depressive symptoms on stressors are not adjusted for shared variance with other variables. In other words, the S→DS coefficients are smaller because, at least in part, these paths are adjusted for their shared variance with the other stressors and represent each stressor's independent association with depressive symptoms.

Limitations

As we have discussed earlier, there are several aspects of this study that may limit the generalizability of the findings. One is that the sample is not representative of adolescents as a whole. Second, even though this study was designed to address the methodological problems regarding life stressors measurement, the methods we used may not have ruled out other sources of measurement error, such as the possibility that the impact of certain stressors varied considerably for different individuals. It is also possible that our checklist measures did not include all the potential stressful circumstances operating in these adolescents' lives. Such shortcomings would effect the estimates of the S-DS associations. In addition, the fact that all data are self-report increases the possibility that estimates may be biased due to method variance.

It is relevant to note, however, that adopting any particular method for assessing stressful life circumstances does not eliminate concerns regarding measurement validity. There has been considerable debate regarding whether normative-based measures (which specify the normative stressfulness and disruptiveness of events) or appraisal-based measures (which allow participants to indicate the degree of event stressfulness) are more appropriate (Compas, 1987; B. P. Dohrenwend, Shrout, Link, Skodol, & Martin, 1986; Lazarus & Folkman, 1986). Normative measures are criticized because they are subject to event variability within category, whereas appraisal measures are criticized because the reports of event stressfulness could be confounded with existing psychological adjustment.

Directions for Future Research

There are several possible directions for future research. As we discussed earlier, most previous life stressor research among adolescents has focused on the experience of White adolescents, whereas most participants in this study were from ethnic minority groups. Future research should examine whether mood-congruent recall bias operates in a similar way for various subgroups of adolescents. Research is also needed to investigate the extent to which experiencing different types of life stressors leads to depressive symptoms for different subgroups of adolescents after accounting for the influence of mood-congruent cognitive bias.

Another way in which this research could be extended would be to further distinguish between different types of stressors rather than rely on the global measures used in this and many other studies. One possibility would be to distinguish between interpersonal and noninterpersonal stressors, as advocated by Rudolph and Hammen (1999). A second possibility would be to create domain-specific measures of major and minor stressors to determine whether adolescents are more (or less) vulnerable to stressors in a particular domain (e.g., home, school, peers, and/or neighborhood).

In this same vein, new approaches to measuring life stressors could be applied. To date, stressors measures assessed through both checklists and interviews have been computed as the total number of stressors experienced. One alternative would be to estimate continuous latent variables of stressors, which could reduce the measurement error that is likely present in the summed variables currently being used. A second possibility would be to apply a type of classification technique, such as latent class analysis. This would enable researchers to examine whether experiencing a certain set of stressors results in an increased likelihood that an adolescent will become more depressed. It will be important that any new measurement strategy investigate the extent to which stressors measures are subject to mood-congruent recall bias.

Notwithstanding the issues to be addressed in future research, we believe that this study demonstrates a valuable approach to studying the S-DS process. Our findings call attention to the necessity of addressing the potential role of mood-congruent cognitive bias in stressors recall, as well as the value of systematically comparing multiple, competing theoretical models.

References

Achenbach, T. M., & Edelbrock, C. (1987). *Manual for the Youth self-report*. Burlington: Department of Psychiatry, University of Vermont.

Blaney, P. H. (1986). Affect and memory: A review. *Psychological Bulletin, 99*, 229–246.

Bower, G. H. (1981). Mood and memory. *American Psychologist, 36*, 129–148.

Burt, C. E., Cohen, L. H., & Bjorck, J. (1988). Perceived family and environment as a moderator of young adolescents' life stress adjustment. *American Journal of Community Psychology, 16*, 101–122.

Caspi, A., Elder, G. H., Jr., & Bem, D. J. (1987). Moving against the world: Life course patterns of explosive children. *Developmental Psychology, 23*, 308–313.

Cicchetti, D. (1984). The emergence of developmental psychopathology. *Child Development, 55*, 1–7.

Clements, M. (2003). Negative life events, daily hassles, and the development of depressive symptoms: A longitudinal study of urban adolescents. (Doctoral dissertation, New York University, 2003). *Dissertation Abstracts International*. (UMI No. 3105851)

Coddington, R. D. (1972). The significance of life events as etiologic factors in the diseases of children: II. A study of a normal population. *Journal of Psychosomatic Research, 16*, 205–213.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, L. H., Burt, C. E., & Bjorck, J. P. (1987). Life stress and adjustment: Effects of life events experienced by young adolescents and their parents. *Developmental Psychology, 23*, 583–592.

Cole, D. A., Nolen-Hoeksema, S., Girgus, J., & Paul, G. (2006). Stress exposure and stress generation in child and adolescent depression: A latent trait-state-error approach to longitudinal analyses. *Journal of Abnormal Psychology, 115*, 40–51.

Compas, B. E. (1987). Stress and life events during childhood and adolescence. *Clinical Psychology Review, 7*, 275–302.

Compas, B. E., Howell, D. C., Phares, V., Williams, R. A., & Ledoux, N. (1989). Parent and child stress and symptoms: An integrative analysis. *Developmental Psychology, 25*, 550–559.

Compas, B. E., Slavin, L. A., Wagner, B. M., & Vannatta, K. (1986). Relationship of life events and social support with psychological dysfunction among adolescents. *Journal of Youth and Adolescence, 15*, 205–221.

DeLongis, A., Folkman, S., & Lazarus, R. S. (1988). The impact of daily stress on health and mood: Psychological and social resources as mediators. *Journal of Personality and Social Psychology, 54*, 486–495.

Dohrenwend, B. P., Shrout, P. E., Link, B. G., Skodol, A. E., & Martin, J. L. (1986). Overview and initial results from a risk factor study of depression and schizophrenia. In J. E. Barrett (Ed.), *Mental disorders in the community: Progress and challenge* (pp. 184–215). New York: Guilford Press.

Dohrenwend, B. S., & Dohrenwend, B. P. (1974). *Stressful life events: Their nature and effects*. New York: Wiley.

Dohrenwend, B. S., Dohrenwend, B. P., Dodson, M., & Shrout, P. E. (1984). Symptoms, hassles, social supports, and life events: Problem of confounded measures. *Journal of Abnormal Psychology, 93*, 222–230.

DuBois, D. L., Felner, R. D., Brand, S., Adan, A. M., & Evans, E. (1992). A prospective study of life stress, social support, and adaptation in early adolescence. *Child Development, 63*, 542–557.

DuBois, D. L., Felner, R. D., Meares, H., & Krier, M. (1994). Prospective investigation of the effects of socioeconomic disadvantage, life stress, and social support on early adolescent adjustment. *Journal of Abnormal Psychology, 103*, 511–522.

Dubow, E. F., Tisak, J., Causey, D., Hryshko, A., & Reid, G. (1991). A two-year longitudinal study of stressful life events, social support, and social problem-solving skills: Contributions to children's behavioral and academic adjustment. *Child Development, 62*, 583–599.

Felner, R. D., & Felner, T. Y. (1989). Primary prevention programs in the educational context: A transactional-ecological framework and analysis. In L. A. Bond & B. E. Compas (Eds.), *Primary prevention and promotion in the schools* (pp. 13–49). Newbury Park, CA: Sage.

Gersten, J. C., Langner, T. S., Eisenberg, J. G., & Simcha-Fagan, O. (1977). An evaluation of the etiologic role of stressful life-change events in psychological disorders. *Journal of Health and Social Behavior, 18*, 228–244.

Grant, K. E., Compas, B. E., Thurm, A. E., McMahon, S. D., & Gipson, P. Y. (2004). Stressors and child and adolescent psychopathology: Measurement issues and prospective effects. *Journal of Clinical Child and Adolescent Psychology, 33*, 412–425.

Hammen, C. (1991). Generation of stress in the course of unipolar depression. *Journal of Abnormal Psychology, 100*, 555–561.

Hammen, C. (1992a). Cognitive, life stress, and interpersonal approaches to a developmental psychopathology model of depression. *Development and Psychopathology, 4*, 189–206.

Hammen, C. (1992b). Life events and depression: The plot thickens. *American Journal of Community Psychology, 20*, 179–193.

Hammen, C., Burge, D., & Adrian, C. (1991). Timing of mother and child depression in a longitudinal study of children at risk. *Journal of Consulting and Clinical Psychology, 59*, 341–345.

Hoffman, J. P., & Su, S. S. (1997). The conditional effects of stress on delinquency and drug use: A strain theory assessment of sex differences. *Journal of Research in Crime and Delinquency, 34*, 46–48.

Holmes, T. H., & Rahe, R. H. (1967). The Social Readjustment Rating Scale. *Journal of Psychosomatic Research, 11,* 213–218.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 1998,* 424–453.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1–55.

Ingram, R. E., Smith, T. W., & Brehm, S. S. (1983). Depression and information processing: Self-schemata and the encoding of self-referent information. *Journal of Personality and Social Psychology, 45,* 412–420.

Johnson, J. H. (1986). *Life events as stressor in childhood and adolescence* (Vol. 8). Beverly Hills, CA: Sage.

Johnson, J. H., & McCutcheon, S. M. (1980). Assessing life stress in older children and adolescents: Preliminary findings with the Life Events Checklist. In I. G. Sarason & C. D. Spielberger (Eds.), *Stress and anxiety* (Vol. 7, pp. 111–127). Washington, DC: Hemisphere.

Kanner, A. D., Coyne, J. C., Schaefer, C., & Lazarus, R. S. (1981). Comparison of two modes of stress measurement: Daily hassles and uplifts versus major life events. *Journal of Behavioral Medicine, 4,* 1–39.

Kim, J. K., Conger, R. D., Elder, G. H., & Lorenz, F. O. (2003). Reciprocal influences between stressful life events and adolescent internalizing and externalizing problems. *Child Development, 74,* 127–143.

Klaassen, T., Riedel, W. J., Deutz, N. E. P., & van Praag, H. M. (2002). Mood congruent memory bias induced by tryptophan depletion. *Psychological Medicine, 32,* 167–172.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York: The Guilford Press.

Larson, R., & Ham, M. (1993). Stress and "storm and stress" in early adolescence: The relationship of negative events with dysphoric affect. *Developmental Psychology, 29,* 130–140.

Lazarus, R. S., & Folkman, S. (1986). Reply to Deutsch and Green. *American Psychologist, 41,* 715–716.

Lee, S., & Hershberger, S. (1990). A simple rule for generating equivalent models in covariance structure modeling. *Multivariate Behavioral Research, 25,* 313–334.

Masten, A. S., Neeman, J., & Andenas, S. (1994). Life events and adjustment in adolescents: The significance of event independence, desirability, and chronicity. *Journal of Research on Adolescence, 4,* 71–97.

McDowall, J. (1984). Recall of pleasant and unpleasant words in depressed subjects. *Journal of Abnormal Psychology, 93,* 401–407.

Meyer, P. A., Garrison, C. Z., Jackson, K. L., Addy, C. L., McKeown, R. E., & Waller, J. L. (1993). Undesirable life-events and depression in young adolescents. *Journal of Child and Family Studies, 2,* 47–60.

Morales, J. R., & Guerra, N. G. (2006). Effects of multiple context and cumulative stress on urban children's adjustment in elementary school. *Child Development, 77,* 907–923.

Muthén, L. K., & Muthén, B. O. (2006). *Mplus User's Guide* (4th ed.). Los Angeles: Author.

Newcomb, M. D., Huba, G. J., & Bentler, P. M. (1986). Desirability of various life change events among adolescents: Effects of exposure, sex, age, and ethnicity. *Journal of Research in Personality, 20,* 207–227.

Nolen-Hoeksema, S., & Girgus, J. S. (1994). The emergence of gender differences in depression during adolescence. *Psychological Bulletin, 115,* 424–443.

Pianta, R. C., & Egeland, B. (1994). Relation between depressive symptoms and stressful life events in a sample of disadvantaged mothers. *Journal of Consulting and Clinical Psychology, 62,* 1229–1234.

Rabkin, J. G., & Struening, E. L. (1976). Life events, stress, and illness. *Science, 194,* 1013–1020.

Rowlison, R. T., & Felner, R. D. (1988). Major life events, hassles, and adaptation in adolescence: Confounding in the conceptualization and measurement of life stress and adjustment revisited. *Journal of Personality and Social Psychology, 55,* 432–444.

Rudolph, K. D. (2002). Gender differences in emotional responses to interpersonal stress during adolescence. *Journal of Adolescent Health, 30,* 3–13.

Rudolph, K. D., & Hammen, C. (1999). Age and gender as determinants of stress exposure, generation, and reactions in youngsters: A transactional perspective. *Child Development, 70,* 660–677.

Rudolph, K. D., Hammen, C., Burge, D., Lindberg, N., Herzberg, D., & Daley, S. E. (2000). Toward an interpersonal life-stress model of depression: The developmental context of stress generation. *Development and Psychopathology, 12,* 215–234.

Sameroff, A. J., & Mackenzie, M. J. (2003). Research strategies for capturing transactional models of development: The limits of the possible. *Development and Psychopathology, 15,* 613–640.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7,* 147–177.

Seidman, E. (1991). Growing up the hard way: Pathways of urban adolescents. *American Journal of Community Psychology, 19,* 173–205.

Seidman, E., Allen, L., Aber, J. L., Mitchell, C., Feinman, J., Yoshikawa, H., et al. (1995). Development and validation of adolescent-perceived microsystem scales: Social Support, Daily Hassles, and Involvement. *American Journal of Community Psychology, 23,* 355–388.

Shrout, P. E., Link, B. G., Dohrenwend, B. P., Skodol, A. E., Stueve, A., & Mirotznik, J. (1989). Characterizing life events as risk factors for depression: The role of fateful life events. *Journal of Abnormal Psychology, 98,* 460–467.

Towbes, L. C., Cohen, L. H., & Glyshaw, K. (1989). Instrumentality as a life-stress moderator for early versus middle adolescents. *Journal of Personality and Social Psychology, 57,* 109–119.

Wagner, B. M., Compas, B. E., & Howell, D. C. (1988). Daily and major life events: A test of an integrative model of psychosocial stress. *American Journal of Community Psychology, 16,* 189–205.

Manuscripts Accepted for Publication

Arguing on the Computer, Deanna Kuhn, Wendy Goh, Kalypso Iordanou, and David Shaenfield.

Neighborhood Poverty and Early Transition to Sexual Activity in Young Adolescents: A Developmental Ecological Approach, Veronique Dupere, Eric Lacourse, J. Douglas Willms, Tama Leventhal, and Richard E Tremblay.

Infants' Use of Lexical-Category-to-Meaning Links in Object Individuation, D. Geoffrey Hall, Kathleen Corrigall, Mijke Rhemtulla, Eleanor Donegan, and Fei Xu.

Searching and Planning: Young Children's Reasoning About Past and Future Event Sequences, Kerry McColgan and Teresa McCormack.

Young Children's Recognition of the Intentionality of Teaching, Margalit Ziv, Ayelet Solomon, and Douglas Frye.

Changes in the Relation of Self-Efficacy Beliefs and Behaviors Across Development, Pamela E. Davis-Kean, L. Rowell Huesmann, Justin Jager, W. Andrew Collins, John E. Bates, and Jennifer E. Lansford.

Categories Influence Predictions About Individual Consistency, Marjorie Rhodes and Susan A. Gelman.

The Influence of Race and Gender on Children's Conversations and Playmate Choices, Patrick Leman and Virginia Lam.

Where's the Essence? Developmental Shifts in Beliefs About the Nature of Essential Features, George Newman and Frank Keil.

Profiles of Disruptive Behavior Across Early Childhood: Contributions of Frustration Reactivity, Physiological Regulation, and Maternal Behavior, Kathryn A. Degnan, Susan D. Calkins, Susan P. Keane, and Ashley L. Hill.

Development of Emotional and Behavioral Regulation in Children Born Extremely and Very Preterm: Biological and Social Influences, Caron Clark, Lianne Woodward, John Horwood, and Stephanie Moor.

Neural Processing of Eye Gaze and Threat-Related Emotional Facial Expressions in Infancy, Stefanie Hoehl and Tricia Striano.

Perception of Dynamic and Static Audiovisual Sequences in 3- and 4-Month-Old Infants, David Lewkowicz.

Don't Teach Children 2 + 2 Is Equal to 4: Typical Arithmetic Problems Hinder Learning of Mathematical Equivalence, Nicole McNeil.

Development of 3D Object Completion in Infancy, Kasey C. Soska and Scott P. Johnson.

Developmental Trajectories of Sex-typed Behavior in Boys and Girls: A Longitudinal Population Study of Children Aged 2½ to 8 Years, Susan Golombok, John Rust, Karyofyllis Zervoulis, Tim Croudace, Jean Golding, and Melissa Hines.

Gene-Environment Contributions to the Development of Infant Vagal Reactivity: The Interaction of Dopamine and Maternal Sensitivity, Cathi Propper, Ginger A. Moore, W. Roger Mills-Koonce, Carolyn Tucker Halpern, Ashley Hill, Susan D. Calkins, Mary Anna Carbone, and Martha Cox.

Maternal Socialization of Positive Affect: The Impact of Invalidation on Adolescent Emotion Regulation and Depressive Symptomatology, Marie B. H. Yap, Nicholas Allen, and Cecile D. Ladouceur.

Children's Insecure Representations of the Interparental Relationship and Their School Adjustment: The Mediating Role of Attention Difficulties, Patrick T. Davies, Meredith J. Woitach, Marcia A. Winter, and E. Mark Cummings.

Fetal Substance Exposure and Cumulative Environmental Risk in an African American Cohort, Chie Yumoto, Sandra Jacobson, and Joseph Jacobson.

Longitudinal Analysis of Intrinsic Motivation and Competence Beliefs: Is There a Relation Over Time?, Birgit Spinath and Ricarda Steinmayr.

Limitations in Four-Year-Old Children's Sensitivity to the Spacing Among Facial Features, Catherine Mondloch and Kendra Thomson.

Postural and Object-Oriented Experiences Advance Early Reaching, Object Exploration, and Means-End Behavior, Michele A. Lobo and James C. Galloway.

Stability and Change of Moral Disengagement and Its Impact on Aggression and Violence in Late Adolescence, Marinella Paciello, Roberta Fida, Carlo

Tramontano, Catia Lupinetti, and Gian Vittorio Caprara.

Automatic Activation of Adolescents' Peer-Relational Schemas: Evidence From Priming With Facial Identity, Lauri Nummenmaa, Kätlin Peets, and Christina Salmivalli.

Effects of Aggresssion on Achievement: Does Conflict With the Teacher Make It Worse?, Deborah Stipek and Sarah Miles.

Effects of Family Violence on Psychopathology Symptoms in Children Previously Exposed to Maltreatment, Andrea Maikovich, Sara Jaffee, Candice Odgers, and Robert Gallop.

The Family Check-Up With High-Risk Indigent Families: Preventing Problem Behavior by Increasing Parents' Positive Behavior Support in Early Childhood, Thomas Dishion, Daniel Shaw, Arin Connell, Frances Gardner, Chelsea Weaver, and Melvin Wilson.

Children's Cortisol and the Quality of Teacher–Child Relationships in Child Care, Jared A. Lisonbee, Jacquelyn Mize, Amie Lapp Payne, and Douglas A. Granger.

Time Course of Visual Attention in Infant Categorization of Cats Versus Dogs: Evidence for a Head Bias as Revealed Through Eye Tracking, Paul C. Quinn, Matthew M. Doran, Jason E. Reiss, and James E. Hoffman.

Early Histories of School-Aged Children With Attention-Deficit/Hyperactivity Disorder, Irene M. Loe, Maria D. Balestrino, Randall A. Phelps, Marcia Kurs-Lasky, Diego Chaves-Gnecco, Jack L. Paradise, and Heidi M. Feldman.

Direct and Indirect Aggression: A Meta-Analytic Review of Gender Differences, Intercorrelations, and Relations to Maladjustment, Noel A. Card, Brian D. Stucky, Gita M. Sawalani, and Todd D. Little.

Children's Responses to Interparental Conflict: A Meta-Analysis of Their Associations With Child Adjustment, Kimberly A. Rhoades.

The Social Ecology of Adolescent Alcohol Misuse, Susan Ennett, Vangie A. Foshee, Karl E. Bauman, Andrea Hussong, Robert Faris, John Hipp, L i Cai, Robert DuRant, and Heathe Luz McNaughton Reyes.

HOW INFANTS KNOW MINDS

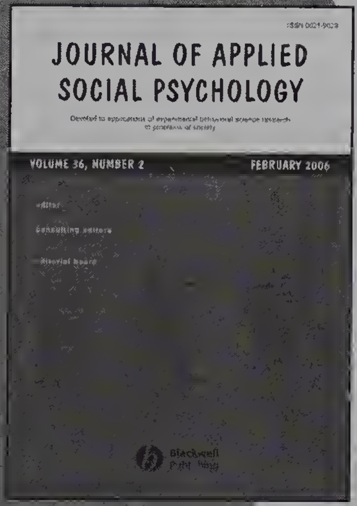
VASUDEVI REDDY

Most psychologists claim that we begin to develop a "theory of mind"—some basic ideas about other people's minds—at age two or three, by inference, deduction, and logical reasoning. But does this mean that small babies are unaware of minds? That they see other people simply as another (rather dynamic and noisy) kind of object? This is a common view in developmental psychology. Yet, as this book explains, there is compelling evidence that babies in the first year of life can tease, pretend, feel self-conscious, and joke with people. Using observations from infants' everyday interactions with their families, Vasudevi Reddy argues that such early emotional engagements show infants' growing awareness of other people's attention, expectations, and intentions.

new in cloth

HOW INFANTS KNOW MINDS
VASUDEVI REDDY

WWW.HUP.HARVARD.EDU **HARVARD UNIVERSITY PRESS**



ISSN 0021-9029

JOURNAL OF APPLIED SOCIAL PSYCHOLOGY

VOLUME 36, NUMBER 2

FEBRUARY 2006

ISSN: 0021-9029

AVAILABLE ONLINE!

JOURNAL OF APPLIED SOCIAL PSYCHOLOGY

Editor: **ANDREW BAUM**

Published since 1971, **JOURNAL OF APPLIED SOCIAL PSYCHOLOGY** is a monthly publication devoted to applications of experimental behavioral science research to problems of society (e.g., organizational and leadership psychology, safety, health, and gender issues; perceptions of war and natural hazards; jury deliberation; performance, AIDS, cancer, heart disease, exercise, and sports).

Blackwell Synergy

JOURNAL OF APPLIED SOCIAL PSYCHOLOGY IS ONLINE AT BLACKWELL SYNERGY! WWW.BLACKWELL-SYNERGY.COM

Sign up to receive Blackwell Synergy free e-mail alerts with complete **JOURNAL OF APPLIED SOCIAL PSYCHOLOGY** tables of contents and quick links to article abstracts from the most current issue. Simply go to www.blackwell-synergy.com, select the journal from the list of journals, and click on ***"Sign-up"*** for FREE e-mail table of contents alerts.

Blackwell Publishing

Call **1-800-835-6770** *(toll free in N. America)*
or: +1 781-388-8599 (US Office); +44 1865 778315 (UK office)
customerservices@blackwellpublishing.com
www.blackwellpublishing.com

View a FREE Online Sample Issue
www.blackwellpublishing.com/jasp

Notice to Contributors

Child Development publishes empirical, theoretical, review, applied, and policy articles reporting research on child development. Published by the interdisciplinary Society for Research in Child Development (SRCD), the journal welcomes relevant submissions from all disciplines. Further information is available at http://www.srcd.org/cd.html.

Types of Articles

Child Development considers manuscripts in formats described below. Inquiries concerning alternative formats should be addressed to the Editor prior to submission. All submissions are expected to be no more than 40 manuscript pages, including tables, references, and figures (but excluding appendices). Authors should provide a justification if the submission is substantially longer. Unless the editor finds that justification compelling, the submission will be returned to the author for shortening prior to editorial review.

Empirical articles comprise the major portion of the journal. To be accepted, empirical articles must be judged as being high in scientific quality, contributing to the empirical base of child development, and having important theoretical, practical, or interdisciplinary implications. Reports of multiple studies, methods, or settings are encouraged, but single-study reports are also considered. Empirical articles will thus vary considerably in length, but should be no longer than 40 manuscript pages; text and graphics should be as concise as material permits. All modes of empirical research are welcome.

Empirical reports are reserved for short cutting-edge empirical papers that are no longer than 4,000 words in length (including text, tables, footnotes, appendices), which advance research and knowledge in an area through noteworthy findings and/or new methods. For manuscripts that require longer descriptions of methods and results, authors should use the Empirical article format.

Reviews focus on past empirical and/or on conceptual and theoretical work. They are expected to synthesize or evaluate a topic or issue relevant to child development, should appeal to a broad audience, and may be followed by a small number of solicited commentaries.

Essays describe original concepts, methods, trends, applications, and theories; these may also be accompanied by solicited commentaries.

Child Development and ... are articles that provide readers with tutorials about some new concept or academic specialty pertinent to research in child development. These papers should review the major definitions, methods, and findings of the concept or specialty and discuss past or potential links to child development.

From another perspective is a format in which papers on a focal topic, written by different authors, are published simultaneously. Papers represent diverse perspectives (e.g., authors whose work represents different populations; different disciplines; different theories, methods, or analytic tools). In some cases, calls for submissions on particular topics will be disseminated through SRCD (via e-mail or SRCD publications), and submissions will undergo normal editorial review. In some cases, a submitted manuscript (e.g., an empirical article) may be selected as a lead article for this format, with invited commentaries providing additional perspectives. The editors also welcome suggestions from readers for topics for this format.

Manuscript Submission

Child Development invites for consideration manuscripts that are neither identical to nor substantially similar to work published or under review elsewhere. In the submission cover letter, please provide details about other published or submitted papers having substantial overlap (including data sets) with the new CD submission to enable editors to judge whether the new submission is sufficiently distinct from other work to warrant consideration. Please note if the paper is posted on a website, see http://www.srcd.org/webposting.html. Editors retain the right to reject manuscripts that do not meet established ethical standards for research or dissemination.

Manuscripts should be submitted electronically to the Child Development Online Submission Site at www.srcd.org/CDsubmit/ as a Word or WordPerfect file. Please also submit a cover letter that contains the name(s) of the author(s) and affiliation(s), and the street address, telephone, fax, and electronic mail address of the corresponding author. A corresponding author's submission to *Child Development* implies that all co-authors have agreed to the content and form of the manuscript and that the ethical standards of SRCD have been followed (see the *Child Development* website or pp. 283–284 of the 2000 SRCD Directory). Any financial interest or conflict of interest must be explained to the Editor in the cover letter. The corresponding author is responsible for informing all coauthors, in a timely manner, of manuscript submission, editorial decisions, reviews, and revisions. Please note that if your submission (or its revision) is accepted for publication, each author must submit a signed Exclusive License Form (http://www.srcd.org/Documents/Publications/CD/cdcopyright.pdf) before the manuscript can enter production. Planning ahead to meet this requirement is especially important for manuscripts with multiple authors.

The manuscript file should be formatted with double-spaced, 12-point type, and should include a single paragraph abstract of 100–120 words. Please follow all guidelines on format, style, and ethics provided in the Publication Manual (5th ed.) of the American Psychological Association. Figures included with initial submissions will not be returned. Therefore, please submit only electronic files or copies of figures. Authors should keep a copy of all correspondence, files, and figures to guard against loss.

Manuscript Review

If you have any questions about your submission, please inquire at cdev@srcd.org or call (734) 926-0615. Each manuscript is handled by the Editor or an Associate Editor who consults with one or more Consulting Editors and/or ad hoc reviewers who have relevant expertise. To ensure blind review, cover sheets are removed before review; authors should avoid including any other information about identity or affiliation in submissions. Copies of the submission and associated correspondence are retained in the SRCD archives. For accepted manuscripts, authors are required to prepare a 300–500 layperson's summary for public dissemination purposes. Details are provided to authors as part of final processing.

There is no charge for publication in *Child Development* unless tabular or graphic materials exceed 10% of the total number of pages. Charges are also levied for changes in proofs other than correction of printer's errors. Any inquiries relating to charges or business matters (including reprint orders) should be addressed to Blackwell Publishing, *Child Development*, Production Coordinator, 350 Main Street, Malden, MA 02148, (781) 388-8200.

Inquiries and suggestions regarding editorial policy may be addressed to: Dr. Jeffrey J. Lockman, Incoming Editor, Department of Psychology, Tulane University, 2007 Percival Stern Hall, New Orleans, LA 70118-5698, lockman@tulane.edu. *Please do not send electronic or paper manuscripts to the above address.* Doing so will delay the processing of your manuscript.

Monographs of the Society for Research in Child Development

Recent

Developing Object Concepts in Infancy: An Associative Learning Perspective—*David H. Rakison and Gary Lupyan* (Serial No. 289, 2008).

The Genetic and Environmental Origins of Learning Abilities and Disabilities in the Early School Years—*Yulia Kovas, Claire M. A. Haworth, Philip S. Dale, and Robert Plomin* (Serial No. 288, 2007).

The Preservation of Two Infant Temperaments Into Adolescence—*Jerome Kagan, Nancy Snidman, Vali Kahn, and Sara Towsley* (Serial No. 287, 2007).

Children's Questions: A Mechanism for Cognitive Development—*Michelle M. Chouinard* (Serial No. 286, 2007).

Best Practices in Quantitative Methods for Developmentalists—*Kathleen McCartney, Margaret R. Burchinal, and Kristen L. Bub* (Serial No. 285, 2006).

Foundations of Self-Awareness: An Exploration Through Autism—*R. Peter Hobson, Gayathri Chidambi, Anthony Lee, and Jessica Meyer* (Serial No. 284, 2006).

Our Children Too: A History of the Black Caucus of the Society for Research in Child Development, 1973–1997—*Diana T. Slaughter-Defoe, Aline M. Garrett, and Algea O. Harrison-Hale* (Serial No. 283, 2006).

Parental Support, Psychological Control, and Behavioral Control: Assessing Relevance Across Time, Culture, and Method—*Brian K. Barber, Heidi E. Stolz, and Joseph A. Olsen* (Serial No. 282, 2005).

Being Hurt and Hurting Others: Children's Narrative Accounts and Moral Judgments of Their Own Interpersonal Conflicts—*Cecilia Wainryb, Beverly A. Brenl, and Sonia Matwin* (Serial No. 281, 2005).

Childhood Sexual Assault Victims: Long-Term Outcomes After Testifying in Criminal Court—*Jodi A. Quas, Gail S. Goodman, Simona Ghetti, Kristen Weede Alexander, Robin Edelstein, Allison D. Redlich, Ingrid M. Cordon, and David P. H. Jones* (Serial No. 280, 2005).

The Emergence of Social Cognition in Three Young Chimpanzees—*Michael Tomasello and Malinda Carpenter* (Serial No. 279, 2005).

Trajectories of Physical Aggression From Toddlerhood to Middle Childhood: Predictors, Correlates, and Outcomes—*NICHD Early Child Care Research Network* (Serial No. 278, 2004).

Constraints on Conceptual Development: A Case Study of the Acquisition of Folkbiological and Folksociological Knowledge in Madagascar—*Rita Astuti, Gregg E. A. Solomon, and Susan Carey* (Serial No. 277, 2004).

Origins and Early Development of Human Body Knowledge—*Virginia Slaughter, Michelle Heron, Linda Jenkins, and Elizabeth Tilse* (Serial No. 276, 2004).

Mother – Child Conversations About Gender: Understanding the Acquisition of Essentialist Beliefs—*Susan A. Gelman, Marianne G. Taylor, and Simone P. Nguyen* (Serial No. 275, 2004).

The Development of Executive Function in Early Childhood—*Philip David Zelazo, Ulrich Müller, Douglas Frye, and Stuart Marcovitch* (Serial No. 274, 2003).

Personal Persistence, Identity Development, and Suicide: A Study of Native and Non-native North American Adolescents—*Michael J. Chandler, Christopher E. Lalonde, Bryan W. Sokol, and Darcy Hallett* (Serial No. 273, 2003).

Personality and Development in Childhood: A Person-Centered Approach—*Daniel Hart, Robert Atkins, and Suzanne Fegley* (Serial No. 272, 2003).

How Children and Adolescents Evaluate Gender and Racial Exclusion—*Melanie Killen, Jennie Lee-Kim, Heidi McGlothlin, and Charles Stangor* (Serial No. 271, 2002).

Child Emotional Security and Interparental Conflict—*Patrick T. Davies, Gordon T. Harold, Marcie C. Goeke-Morey, and E. Mark Cummings* (Serial No. 270, 2002).

The Developmental Course of Gender Differentiation: Conceptualizing, Measuring, and Evaluating Constructs and Pathways—*Lynn S. Liben and Rebecca S. Bigler* (Serial No. 269, 2002).

The Development of Mental Processing: Efficiency, Working Memory, and Thinking—*Andreas Demetriou, Constantinos Christou, George Spanoudis, and Maria Platsidou* (Serial No. 268, 2002).

SRCD
Society for Research
in Child Development

WILEY-BLACKWELL

Order from Blackwell Publishing at subscrip@bos.blackwellpublishing.com

This journal is available online. Contact your librarian or visit www.blackwell-synergy.com.

Child Development

ume 79 Number 5
tember/October 2008

Editor
Lynn S. Liben

Associate Editors
Joseph P. Allen
Neil E. Berthier
Marvin W. Daehler
Nancy E. Hill
Kathy A. Hirsh-Pasek
Campbell Leaper
Patricia H. Miller
Sharon Nelson-Le Gall
Paul C. Quinn
Theresa A. Thorkildsen

CURRENT YR/VOL

Marygrove College Library
8425 West McNichols Road
Detroit, MI 48221

Published by Wiley-Blackwell for the Society for Research in Child Development

Published by Wiley-Blackwell for the Society for Research in Child Development

Child Development

CHILD DEVELOPMENT (ISSN 0009-3920), one of three publications of the Society for Research in Child Development, Inc., is published bimonthly by Wiley Subscription Services, Inc., a Wiley Company, 111 River St., Hoboken, NJ 07030-5774. Periodical Postage Paid at Hoboken, NJ and additional offices.

Postmaster: Send all address changes to *CHILD DEVELOPMENT*, Journal Customer Services, John Wiley & Sons Inc., 350 Main Street, Malden, MA 02148-5020.

Information for Subscribers: *Child Development* is published in six issues per year. Subscription prices for 2008 are: Premium Institutional: US$585 (The Americas), £417 (Rest of World). Premium Online-Only rates are also available (see below). Prices are exclusive of tax. Australian GST, Canadian GST and European VAT will be applied at the appropriate rates. For more information on current tax rates, please go to www.wiley.com, click on Help and follow the link through to Journal subscriptions. The Premium institutional price includes online access to the current and all online back files to January 1st 1997, where available. For other pricing options, including access information and terms and conditions, please visit www.interscience.wiley.com/journals.

Delivery Terms and Legal Title: Prices include delivery of print journals to the recipient's address. Delivery terms are Delivered Duty Unpaid (DDU); the recipient is responsible for paying any import duty or taxes. Legal title passes to the customer on despatch by our distributors.

For more information about Wiley-Blackwell journals, including online access information, terms and conditions, and other pricing options, please visit www.blackwellpublishing.com or contact our customer service department, tel: (800) 835-6770 or +1 (781) 388-8598 (US office); +44 (0) 1865 778315 (UK office); or +65 6511 8000 (Asia Office). Subscription in 2008 includes three issues of *Monographs of the Society for Research in Child Development* and three issues of *Child Development Perspectives*. Single issues from current and recent volumes are available at the current single issue price from Wiley-Blackwell Journals. Earlier issues may be obtained from Periodicals Service Company, 11 Main Street, Germantown, NY 12526, USA. Tel: +1 518 537 4700, Fax: +1 518 537 5899, Email: psc@periodicals.com. Volumes 37 to date are available in microfilm from ProQuest Information and Learning, Serials Customer Service Department, 300 North Zeeb Road, Ann Arbor, MI 48106; telephone: (800) 521-0600 x2873.

Electronic Access: The *Child Development* website can be found at www.srcd.org/cd.html. Abstract information for this journal is electronically available at www.blackwellpublishing.com/cdev. This journal is available online. Visit www.blackwell-synergy.com to search the articles and register for table of contents e-mail alerts.

For new orders, renewals, sample copy requests, claims, changes of address and all other subscription correspondences please contact the Journals Department at your nearest Wiley-Blackwell office (address details listed above). UK office phone: +44 (0) 1865-778315, Fax: +44 (0) 1865-471775, Email: cs-journals@wiley.com; US office phone: 800-835-6770 or 781-388-8598, Fax: 781-388-8232, Email: cs-journals@wiley.com; Asia office phone: +65 6511 8000, Fax: +44 (0) 1865-471775, Email: cs-journals@wiley.com.

Membership communication and address changes should be addressed to the SRCD Executive Office, 2950 South State Street, Suite 401, Ann Arbor, MI 48104, USA. http://www.srcd.org/

Copyright and Photocopying: Journal Compilation © 2008 The Society for Research in Child Development, Inc. All rights reserved. No part of this publication (apart from articles marked "OnlineOpen") may be reproduced, stored or transmitted in any form or by any means without the prior permission in writing from the copyright holder. Authorization to photocopy items for internal and personal use is granted by the copyright holder for libraries and other users registered with their local Reproduction Rights Organization (RRO), e.g., Copyright Clearance Center (CCC), 222 Rosewood Drive, Danvers, MA 01923, USA (www.copyright.com), provided the appropriate fee is paid directly to the RRO. This consent does not extend to other kinds of copying such as copying for general distribution, for advertising and promotional purposes, for creating new collective works or for resale. OnlineOpen articles are made available in accordance with the terms of the Creative Commons Deed, Attribution 2.5 (further details available from www.creativecommons.org), which allows Open Access dissemination of the article, but does not permit commercial exploitation or the creation of derivative works. Special requests should be addressed to jrights@wiley.com.

For advertising information, please visit the journal's website at www.blackwellpublishing.com/cdev or contact the Academic and Science Advertising Sales Coordinator at journaladsUSA@bos.blackwellpublishing.com. 350 Main St., Malden, MA 02148. Phone: 781.388.8532, Fax: 781.338.8532.

Indexing/Abstracting: The contents of this journal are indexed or abstracted in the following: *Academic ASAP; Expanded Academic ASAP; Academic Search Elite; Academic Search Premier; Applied Social Sciences Index and Abstracts; Biomedical Reference Collection: Comprehensive Edition; CAB Abstracts; CABDirect; Contents Pages in Education; Current Contents/Social & Behavioral Science; Current Index to Journals in Education; Cambridge Scientific Abstracts Linguistics & Language Behavior Abstracts; CatchWord; Current Index to Journals in Education; Educational Research Abstracts; Environmental Sciences & Pollution Management; e-psyche; EBSCO Sociological Collection; EBSCO Online; EBSCO Professional Development Collection; Educational Research Abstracts; Educational Resources Information Center; Health and Safety Science Abstracts; ISI Basic Social Sciences Index; InfoTrac OneFile; Ingenta; JCR Social Sciences Edition; Leisure, Recreation, and Tourism Abstracts; Online Computer Library Center FirstSearch Electronic Collections Online; Periodical Abstracts Library; Periodical Abstracts Research II; ProQuest 5000; ProQuest Social Science Plus Text; PsycINFO; Psychological Abstracts; Psychology and Behavioral Sciences Collection; Pollution Abstracts; Safety Science & Risk Abstracts; Sage Family Studies Abstracts; Social Services Abstracts; Social Science Citation Index; Sociological Abstracts; Student Resource Center Bronze; Student Resource Center Gold; Student Resource Center Silver; Violence & Abuse Abstracts; Wilson OmniFile V; Wilson Social Sciences Abstracts;* and *Wilson Social Sciences Index.*

The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials, ANSI Z39.48—1984. This journal is printed on acid-free paper.

Published by Wiley-Blackwell for the Society for Research in Child Development

Child Development

Editor
Lynn S. Liben
The Pennsylvania State University

Associate Editors

Joseph P. Allen, *University of Virginia*
Neil E. Berthier, *University of Massachusetts*
Marvin W. Daehler, *University of Massachusetts*
Nancy E. Hill, *Duke University*
Kathy A. Hirsh-Pasek, *Temple University*
Campbell Leaper, *University of California, Santa Cruz*
Patricia H. Miller, *The University of Georgia*
Sharon Nelson-Le Gall, *University of Pittsburgh*
Paul C. Quinn, *University of Delaware*
Theresa A. Thorkildsen, *University of Illinois at Chicago*

Managing Editor
Detra N. Davis

Editorial Assistant
Kaitlin MacKenzie

Consulting Editors

Brian P. Ackerman
Daisuke Akiba
Eric M. Anderman
Roger Bakeman
Lynne E. Baker-Ward
Karen Bartsch
Amy D. Bellmore
Ramesh S. Bhatt
Rebecca S. Bigler
Pamela Blewitt
Celia A. Brownell
Peter Bryant
Kristin Buss
Susan D. Calkins
Tara C. Callaghan
Joseph J. Campos
Deborah M. Capaldi
Noel A. Card
Zhe Chen
Pamela M. Cole
John Colombo
William E. Cross Jr.
Pamela E. Davis-Kean
Adele Diamond
Mary Ann Foley
Bridget Franks
Andrew J. Fuligni
Mary Gauvain
Susan L. Golbeck
H. Hill Goldsmith
Roberta M. Golinkoff
Douglas Granger
Catherine A. Haden
Laura D. Hanish
Charles C. Helwig
Gail D. Heyman
Erika Hoff
George J. Hollich
Alice Sterling Honig
Cynthia Hudley
Jennifer Jenkins
Susan Scanlon Jones
Deborah G. Kemler Nelson
Melanie Killen
Roger Kobak
Grazyna Kochanska
Brett Laursen
Bonnie Jean Leadbeater
Elizabeth Ann Lemerise
Yonata Levy
Joan M. Lucariello
Sarah C. Mangelsdorf
Denis Mareschal
Herb W. Marsh
Clay W. Mash
Susan M. McHale
Derek E. Montgomery
Ulrich Mueller
Tamera Murdock
Letitia R. Naigles
Ageliki Nicolopoulou
Thomas G. O'Connor
Yukari Okamoto
Jamie M. Ostrov
David G. Perry
Jodie M. Plumert
Gary Resnick
Lori A. Roggman
Glenn I. Roisman
Amanda J. Rose
Kenneth H. Rubin
Diane N. Ruble
Martin D. Ruck
Ted Ruffman
Dale H. Schunk
Margaret L. Signorella
Judith G. Smetana
David Sobel
Christia Spears Brown
Cynthia A. Stifter
Mike L. Stoolmiller
Susan Sy
Lisa E. Szechter
Rebecca Treiman
Georgene L. Troseth
Marion K. Underwood
Marinus H. van IJzendoorn
Brian E. Vaughn
Melissa K. Welch-Ross
Allan Wigfield
Terry Wood
Jacqueline D. Woolley
Barbara Younger-Rossmann

Child Development, September/October 2008, Volume 79, Number 5, Pages 1185–1595

Contents

EMPIRICAL ARTICLES

Direct and Indirect Aggression During Childhood and Adolescence: A Meta-Analytic Review of Gender Differences, Intercorrelations, and Relations to Maladjustment
Noel A. Card, Brian D. Stucky, Gita M. Sawalani, and Todd D. Little 1185

Development of Three-Dimensional Object Completion in Infancy
Kasey C. Soska and Scott P. Johnson 1230

Young Children's Recognition of the Intentionality of Teaching
Margalit Ziv, Ayelet Solomon, and Douglas Frye 1237

Changes in the Relation of Self-Efficacy Beliefs and Behaviors Across Development
Pamela E. Davis-Kean, L. Rowell Huesmann, Justin Jager, W. Andrew Collins, John E. Bates, and Jennifer E. Lansford 1257

Categories Influence Predictions About Individual Consistency
Marjorie Rhodes and Susan A. Gelman 1270

Stability and Change of Moral Disengagement and Its Impact on Aggression and Violence in Late Adolescence
Marinella Paciello, Roberta Fida, Carlo Tramontano, Catia Lupinetti, and Gian Vittorio Caprara 1288

Arguing on the Computer: A Microgenetic Study of Developing Argument Skills in a Computer-Supported Environment
Deanna Kuhn, Wendy Goh, Kalypso Iordanou, and David Shaenfield 1310

The Influence of Race and Gender on Children's Conversations and Playmate Choices
Patrick J. Leman and Virginia L. Lam 1329

Where Is the Essence? Developmental Shifts in Children's Beliefs About Internal Features
George E. Newman and Frank C. Keil 1344

Profiles of Disruptive Behavior Across Early Childhood: Contributions of Frustration Reactivity, Physiological Regulation, and Maternal Behavior
Kathryn A. Degnan, Susan D. Calkins, Susan P. Keane, and Ashley L. Hill-Soderlund 1357

Gene–Environment Contributions to the Development of Infant Vagal Reactivity: The Interaction of Dopamine and Maternal Sensitivity
Cathi Propper, Ginger A. Moore, W. Roger Mills-Koonce, Carolyn Tucker Halpern, Ashley L. Hill-Soderlund, Susan D. Calkins, Mary Anna Carbone, and Martha Cox 1377

The Family Check-Up With High-Risk Indigent Families: Preventing Problem Behavior by Increasing Parents' Positive Behavior Support in Early Childhood
Thomas J. Dishion, Daniel Shaw, Arin Connell, Frances Gardner, Chelsea Weaver, and Melvin Wilson 1395

Maternal Socialization of Positive Affect: The Impact of Invalidation on Adolescent Emotion Regulation and Depressive Symptomatology
Marie B. H. Yap, Nicholas B. Allen, and Cecile D. Ladouceur 1415

Infants' Use of Lexical-Category-to-Meaning Links in Object Individuation
D. Geoffrey Hall, Kathleen Corrigall, Mijke Rhemtulla, Eleanor Donegan, and Fei Xu 1432

Development of Emotional and Behavioral Regulation in Children Born Extremely Preterm and Very Preterm: Biological and Social Influences
Caron A. C. Clark, Lianne J. Woodward, L. John Horwood, and Stephanie Moor 1444

Neighborhood Poverty and Early Transition to Sexual Activity in Young Adolescents: A Developmental Ecological Approach
Véronique Dupéré, Éric Lacourse, J. Douglas Willms, Tama Leventhal, and Richard E. Tremblay 1463

Searching and Planning: Young Children's Reasoning About Past and Future Event Sequences
Kerry L. McColgan and Teresa McCormack 1477

Effects of Family Violence on Psychopathology Symptoms in Children Previously Exposed to Maltreatment
Andrea Kohn Maikovich, Sara R. Jaffee, Candice L. Odgers, and Robert Gallop 1498

Limitations in 4-Year-Old Children's Sensitivity to the Spacing Among Facial Features
Catherine J. Mondloch and Kendra Thomson 1513

Limitations to Teaching Children 2 + 2 = 4: Typical Arithmetic Problems Can Hinder Learning of Mathematical Equivalence
Nicole M. McNeil 1524

Perception of Dynamic and Static Audiovisual Sequences in 3- and 4-Month-Old Infants
David J. Lewkowicz 1538

Longitudinal Analysis of Intrinsic Motivation and Competence Beliefs: Is There a Relation Over Time?
Birgit Spinath and Ricarda Steinmayr 1555

Children's Insecure Representations of the Interparental Relationship and Their School Adjustment: The Mediating Role of Attention Difficulties
Patrick T. Davies, Meredith J. Woitach, Marcia A. Winter, and E. Mark Cummings 1570

Developmental Trajectories of Sex-Typed Behavior in Boys and Girls: A Longitudinal General Population Study of Children Aged 2.5–8 Years
Susan Golombok, John Rust, Karyofyllis Zervoulis, Tim Croudace, Jean Golding, and Melissa Hines 1583

OTHER

Manuscripts Accepted 1594

Child Development, September/October 2008, Volume 79, Number 5, Pages 1185–1229

Direct and Indirect Aggression During Childhood and Adolescence: A Meta-Analytic Review of Gender Differences, Intercorrelations, and Relations to Maladjustment

Noel A. Card
University of Arizona

Brian D. Stucky
University of North Carolina Chapel Hill

Gita M. Sawalani
University of Kansas

Todd D. Little
University of Kansas

This meta-analytic review of 148 studies on child and adolescent direct and indirect aggression examined the magnitude of gender differences, intercorrelations between forms, and associations with maladjustment. Results confirmed prior findings of gender differences (favoring boys) in direct aggression and trivial gender differences in indirect aggression. Results also indicated a substantial intercorrelation ($\bar{r} = .76$) between these forms. Despite this high intercorrelation, the 2 forms showed unique associations with maladjustment: Direct aggression is more strongly related to externalizing problems, poor peer relations, and low prosocial behavior, and indirect aggression is related to internalizing problems and *higher* prosocial behavior. Moderation of these effect sizes by method of assessment, age, gender, and several additional variables were systematically investigated.

Child and adolescent aggression has received longstanding theoretical and empirical attention. The reason for this attention is clear: Aggressive behaviors, on average, are associated with maladjustment for both aggressors (Coie & Dodge, 1998; S. Feshbach, 1970; Ladd, 2005; Parke & Slaby, 1983; cf. Hawley, Little, & Rodkin, 2007) and victims (Card, 2003; Card, Isaacs, & Hodges, 2007). Historically, research has focused on direct forms of aggression, especially acts of physical aggression. Because boys typically enact more physical aggression than girls (Hyde, 1984; Maccoby & Jacklin, 1974), less research has focused on girls' aggression. Some studies have not included girls in their samples, and even studies that included boys and girls have not always analyzed potential gender differences (Björkqvist, 1994).

In the 1980s and 1990s (e.g., Björkqvist, Lagerspetz, & Kaukiainen, 1992; Cairns, Cairns, Neckerman, Ferguson, & Gariépy, 1989; Crick, 1995; Crick & Grotpeter, 1995; Lagerspetz, Björkqvist, & Peltonen, 1988; Lancelotta & Vaughn, 1989; for earlier work, see Buss, 1961; N. D. Feshbach, 1969), researchers began considering a wider range of aggressive behaviors, including forms of aggression that are more covert or indirect in nature and aimed at damaging the target's social relations (rather than, e.g., inflicting bodily harm). These forms of aggression have been given various names, including indirect, covert, relational, and social aggression. The term *indirect aggression* was introduced by N. D. Feshbach (1969) to refer to behaviors harming a target by rejection or exclusion. Lagerspetz et al. (1988) later used this term to refer to behaviors such as gossiping, befriending others, and exclusion that do not directly confront the victim. Shortly thereafter, the term *social aggression* was used by Cairns et al. (1989) to refer to children's descriptions of manipulations in group acceptance through ostracism or character attacks. Galen and Underwood (1997) used this same term to refer to behaviors such as rejection, negative nonverbal expressions, rumor spreading, or social exclusion aimed at damaging the victim's self-esteem or social status. Crick (e.g., 1995) used the term *relational aggression* to refer to harming others through manipulation of peer relationships (e.g., threatening to terminate friendship, excluding from group).

We thank Juan Casas and Jamie Ostrov for commenting on a previous draft of this work. This work was supported in part by a National Institute of Mental Health Individual National Research Service Award (F32 MH072005) to the first author, a University of Kansas Undergraduate Research Award to the second author, and a NFGRF grant (2301779) from the University of Kansas to the fourth author.

Correspondence concerning this article should be addressed to Noel A. Card, Division of Family Studies and Human Development, University of Arizona, 1110 E. South Campus Drive, Tucson, AZ 85721-0033. Electronic mail may be sent to: ncard@email.arizona.edu.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7905-0001

When precisely applied, each term refers to slightly different behaviors (Xie, Cairns, & Cairns, 2005). Although the existing research lacks perfect correspondence between terminology used and operational definitions employed, the constructs converge around the common theme of behaviors that attack a victim's actual or perceived social relations with others, often (though not always) in a way that avoids direct confrontation. In practice, most researchers have accepted some blurring at the boundaries among these definitions (Archer & Coyne, 2005; Björkqvist, 2001; Underwood, Galen, & Paquette, 2001). This practice is supported by factor-analytic studies indicating two forms of aggressive behavior (e.g., Crick & Grotpeter, 1995; Grotpeter & Crick, 1996; Hart, Nelson, Robinson, Olsen, & McNeilly-Choque, 1998; Vaillancourt, Brendgen, Boivin, & Tremblay, 2003). One factor includes physical reacts such as hitting, pushing, and tripping as well as overt verbal attacks such as name calling, taunting, and threatening. The second factor includes hurtful manipulation of relationships (e.g., sabotaging the target's friendships or romantic relationships) and damaging the target's social position (e.g., spreading gossip, excluding from activities), often through indirect or covert means. In this article, we refer to these two broad forms of aggression as direct and indirect aggression, respectively, based on both historical precedence (N. D. Feshbach, 1969; Lagerspetz et al., 1988) and arguments supporting this terminology (Archer & Coyne, 2005; Björkqvist, 2001). The term *indirect aggression* is also limited in that it excludes more direct attacks on social well-being, but it is a reasonable representation of this heterogeneous class of behaviors. Therefore, we adapt an inclusive definition of these two forms, including both physical and verbal aggression within direct aggression and the various terms social, relational, covert, and indirect aggression within our consideration of indirect aggression.

Gender Differences in Direct and Indirect Aggression

Attention to these indirect forms has led to a resurgence of interest in potential gender differences in aggression. Some have argued that although boys enact more direct aggression than girls, the indirect forms of aggression may be more typical of girls than of boys (e.g., Björkqvist et al., 1992; Crick, 1997). Based on etiological explanations for aggressive behaviors, several explanations for why girls might be more indirectly aggressive than boys have been offered (Archer, 2004). One is based on sexual selection theory, suggesting that attacks aimed at girls' sexual reputations (e.g., gossip about promiscuity) and their social reputations are especially prominent in intrafemale competition (Artz, 2005). A related biological explanation is that females' lower physical strength necessitates girls' reliance on indirect means of aggression more so than boys (Björkqvist, 1994). Another explanation is based on findings that girls' peer groups are characterized by fewer, closer friendships relative to boys' more numerous, casual relationships (Maccoby, 1990, 1998), suggesting that indirect aggression is particularly hurtful among girls because it targets these relationships (Crick, Bigbee, & Howes, 1996; Galen & Underwood, 1997). Finally, explanations emphasizing socialization argue that differences in the degree to which parents and other adults discourage directly aggressive behavior of girls versus boys leads girls to use more covert forms of aggression (Underwood, 2003). Such differential socialization might further occur within children's sex-segregated peer groups (Maccoby, 1990, 1998), which provides models and reinforcements for appropriate forms of enacting aggression (e.g., indirect aggression within girls' peer groups).

Despite conceptual expectations that girls enact more indirect aggression than boys, the existing literature documenting these gender differences is inconsistent. Archer (2004) performed a meta-analytic review of 78 studies that compared indirect forms of aggression enacted by males versus females from childhood throughout adulthood. His results indicated that, averaged across age, females exhibit more indirect aggression than males when assessed using some methods (i.e., observations, peer ratings, teacher reports) but not others (e.g., peer nominations, self-reports), with the magnitude of these gender differences ranging from small to medium. Archer's results (by method of assessment) specific to childhood and adolescence did not indicate gender differences in indirect aggression, although the simultaneous disaggregation by reporting method and age may have limited the ability to detect such differences.

The first goal of the current meta-analytic review was to examine the magnitude of gender differences in both direct and indirect aggression among children and adolescents. Although this first goal partially overlaps with the previous meta-analysis by Archer (2004), the current meta-analysis extends our understanding in several ways. First, we focus more explicitly on childhood and adolescence. Second, we include more studies during this developmental period (as described below, our examination of gender differences includes 107 studies of samples younger than 18 years, whereas Archer's contained 52 studies assessing this age range), allowing for more precise results and stronger generalizations.

Third, we consider a wider range of methods in assessing aggression. Fourth, we analyze corrected effect sizes (see below) to gain a more accurate representation of true gender differences. Finally, we use random-effects models that allow greater generalization of findings than the fixed-effects models used by Archer.

In addition to examining the magnitude of gender differences in direct and indirect aggression, we evaluate several potential moderators of these differences. First, we consider age moderation. Although Archer (2004) failed to find evidence of age moderation, we expect gender differences to increase between childhood and adolescence due to (a) greater engagement in gender-typical behavior; (b) increasing time spent with peers, especially same-sex peers; (c) biological changes in puberty (e.g., greater gender differentiation in physical stature); and (d) possible increases in intrasex competition (Trivers, 1972; cf. Bjorklund & Pellegrini, 2000; Wilson & Daly, 1985). We also evaluate method of assessing aggression as a potential moderator. Expectations regarding this moderator are less clear. However, we hypothesize that adults' expectations for gender-stereotypic behavior would result in teachers and parents perceiving girls as more indirectly aggressive than boys more so than peer reports, self-reports, or observations. Although we do not have specific hypotheses, we also evaluate several additional moderators of gender differences. We consider two aspects of the context from which samples were drawn: the proportion of ethnic minorities in the sample and the country in which the study was conducted. We also evaluate moderation by the gender of the first author of the study in order to explore whether gender-linked expectations influence findings regarding gender differences. In addition, we consider whether a focus on gender in the study, operationalized by whether the term *gender* (or similar terms) appeared in the title. Finally, we consider publication status in order to evaluate potential biases in the results.

Intercorrelation Between Direct and Indirect Aggression

Although factor-analytic studies have supported the distinctiveness of indirect forms of aggression from direct forms, even a casual perusal of this literature suggests that these two forms of aggression are often highly intercorrelated. The second goal of this meta-analytic review, then, was to evaluate the magnitude of the correlation between direct and indirect aggression and to examine potential moderators of this interrelation. The magnitude of the correlation between these two forms speaks to the important issue of whether they should be viewed as different manifestations of a common underlying construct (if correlations are high) or as separate, independent constructs (if correlations are low). The magnitude of this correlation also speaks to, albeit indirectly, the potential origins of these forms: High correlations would imply common origins such as shared genetic contributions, common evolutionary functions, or common parental or peer socialization for the two forms. In contrast, modest correlations suggest differences in the origins of these two forms. Although our meta-analysis does not attempt to identify these shared versus unique antecedents, the magnitude of the association between direct and indirect forms of aggression can suggest in which direction we should look.

We consider several potential moderators of these correlations, although little prior evidence is available to suggest directions of these moderators. Against the viable alternative of no change, the two forms of aggression may become either more differentiated or more overlapping with age. On the one hand, several reasons for expecting decreasing correlations between direct and indirect aggression with age can be offered. With age, understanding of the outcomes of aggression (both rewards and consequences) and cognitive capacity to enact various forms of aggression may increase. Peer group pressure for "gender normative" forms of aggression may also increase with age. Increasing biological differences between boys and girls following puberty and increasing saliency of intrasexual competition (and the use of specific forms of aggression to achieve dominance) might also predict greater differentiation with age. Together, these possibilities suggest that older children and adolescents are more likely to enact a "preferred" form of aggression than would younger children, thus leading to a decrease in the magnitude of correlation between direct and indirect aggression with age (i.e., divergence). On the other hand, verbal skills and understanding of social dynamics increase with age (combined with the maintained capacity to enact physical aggression) making all forms available among older children. With increasing age, individuals with a propensity to behave aggressively (e.g., if a common genetic origin for aggressiveness exists) would increasingly use all available forms. This view suggests higher correlations between direct and indirect forms at higher ages (i.e., convergence). Because either possibility is plausible but prior research has not illuminated this issue, we investigate age moderation of this correlation.

Similarly, the prior literature provides little guidance to indicate whether measuring direct and indirect aggression via some sources results in different magnitudes of overlap than do other sources. Such differences, however, might be expected. For example, because trained observers should distinguish between these two forms precisely, observational measures may yield lower correlations between direct and indirect aggression than studies using other sources. In this meta-analysis, we compare the magnitudes of the correlations between direct and indirect aggression when assessed via self-, peer, teacher, and parent reports and observations.

We also examine gender moderation in the magnitude of the correlation between direct and indirect aggression. The direction of this potential moderator is likely to depend on the magnitude of any gender differences (see above). For example, if girls enact more indirect aggression than boys (whereas boys enact more direct aggression than girls), then the correlation between direct and indirect aggression may be similar for boys and girls (i.e., boys enact primarily direct aggression with some secondary indirect aggression, whereas girls enact primarily indirect aggression with some secondary direct aggression). If a gender difference in indirect aggression is not present, then these forms may be more highly intercorrelated among boys (for whom both forms are viable options) than girls (for whom direct aggression is a less viable option and therefore likely enacted by different individuals).

Finally, we systematically explore potential moderation of this intercorrelation by ethnic and national context, author gender, whether the study focused on gender, and publication status.

Associations of Direct and Indirect Aggression With Maladjustment

As noted, aggressive behaviors, on average, are associated with maladjustment. However, far less is known about the specific links that may exist between maladjustment and direct versus indirect forms of aggression. Research emerging in the past couple decades has examined these associations, but the current meta-analysis is the first to quantitatively summarize the associations of direct and indirect aggression with four indices of adjustment: internalizing problems, externalizing problems, prosocial behavior (an index of positive adjustment), and peer relations. Although these constructs represent a small sample of adjustment indices, they give a general indication of the psychosocial adjustment of directly and indirectly aggressive children.

An important question to consider regarding the distinction between direct and indirect aggression is whether these two forms *differentially* relate to maladjustment. Therefore, we examine the independent and differential magnitudes of association (defined below) of direct and indirect aggression with these four indices of adjustment to identify specific links. We expect that direct forms of aggression, which are often considered less socially acceptable than indirect forms of aggression, are more strongly linked to externalizing behavior problems, low prosocial behaviors, and poorer peer relations. In contrast, we expect indirect aggression to be more strongly linked to internalizing problems than direct aggression. The reason for this latter expectation is that aggression in general shares some antecedents with internalizing problems, such as aversive home environments, problematic peer relationships, and certain social cognitions (e.g., negative view of others). We expect that these potential links between internalizing problems and aggression would manifest primarily in the form of indirect aggression because children who suffer these internalizing problems may prefer the less risky, more covert nature of indirect forms if they are inclined to act aggressively.

The associations of direct and indirect aggression to maladjustment may also vary by age and gender. Although the empirical evidence for developmental trends is scarce (Baillargeon et al., 2007; Tremblay, 2000), some (e.g., Björkqvist, 1994) have suggested that direct aggression peaks during early childhood (this tendency actually differs for physical vs. verbal aspects of direct aggression, with the former peaking earlier than the latter). In contrast, Björkqvist (1994) suggested that indirect aggression peaks during later childhood and adolescence, paralleling the development of greater verbal abilities and social sophistication. Although this potential pattern does not imply that direct aggression is absent during adolescence or that indirect forms of aggression are not enacted by young children, it does suggests two possible directions of age moderation. First, engaging in forms of aggression that are developmentally less appropriate (i.e., direct aggression during adolescence and indirect aggression during early childhood) may be associated with greater maladjustment, perhaps because this aggression elicits harsher consequences (e.g., more severe consequences by parents and teachers and more rejection by peers). Alternatively, aggression that is developmentally normative may reflect a typical manifestation of maladjustment and therefore may be more strongly linked to maladjustment.

Gender is another potential moderator of the associations of direct and indirect aggression with adjustment. Our expectations of the direction of moderation are rooted in the concept of gender normative forms of aggression, namely, that direct aggression is normative for boys and indirect aggression is normative for girls (Crick, 1997). This concept does not imply that either form is used exclusively by either gender nor does it specify the forces that might differentially promote or inhibit these forms among one gender or the other. Nevertheless, the gender normative concept suggests two potential directions of gender moderation. First, engaging in gender nonnormative forms of aggression represents greater maladjustment and leads to greater social censure by adults or peers than does engaging in gender normative forms (Crick, 1997). This possibility would result in direct (more so than indirect) aggression being related to maladjustment among girls, whereas indirect (more so than direct) aggression would be related to maladjustment among boys. A second possibility is predicated on the existence of greater variability in gender normative than gender nonnormative aggression. This difference could lead to indirect aggression relating to maladjustment more strongly among girls, whereas boys may exhibit a pattern of direct aggression being more strongly related to maladjustment.

In addition to age and gender, we also systematically explore other moderators of these associations between the two forms of aggression and maladjustment. Specifically, we consider method of assessing aggression, sample ethnicity and nationality, author gender, whether the study title indicated a focus on gender, and publication status.

Summary of Goals

To summarize, we have three primary goals. First, we seek to examine potential gender differences in direct and indirect aggression during childhood and adolescence, a goal that overlaps with but expands upon prior reviews (Archer, 2004; Hyde, 1984; Maccoby & Jacklin, 1974). Second, we examine the magnitude of intercorrelation between direct and indirect aggression in order to determine the extent that these represent common manifestations of a single behavior (i.e., aggressiveness) versus distinct constructs. Third, we investigate the associations of direct and indirect aggression with several aspects of maladjustment, aiming especially to identify the unique, differential links each form has with maladjustment. As a secondary goal, we systematically investigate several potential moderators of these effect sizes.

Method

Selection of Studies

We obtained studies through five approaches. First, we conducted computerized literature searches of the PsycINFO, ERIC, and Proquest (for unpublished dissertations) databases using the keywords "indirect aggression," "relational aggression," "social aggression," or "covert aggression" (last search performed August 2007; we did not include terms specific to direct aggression because we wanted to identify studies with both forms, as described below). Second, we examined references cited in other articles (i.e., "backward search" procedures). Third, we asked several experts in this area to examine earlier lists of reports and suggest other studies. Fourth, we e-mailed all first authors and dissertation supervisors of otherwise obtained reports to request additional studies. Finally, we requested additional studies through the Society for Research in Child Development (SRCD) Peer Relations listserve. The latter two approaches were designed to obtain unpublished studies that might not have been obtained through other means. We chose to be lenient in our initial review of studies in order to be as inclusive as possible, so we reviewed a total of 588 journal articles, chapters, books, dissertations, and unpublished studies in selecting studies for potential inclusion.

We included studies that met three criteria. First, they presented data on both direct and indirect forms of aggression that were relevant to any of the questions of interest in this meta-analysis (i.e., gender differences, intercorrelations, associations with maladjustment). Second, the sample consisted of children younger than 18 years (i.e., adult samples were excluded). Third, the sample could be considered normative (as opposed to including only children in psychiatric or criminal settings). Using these criteria, a total of 148 independent studies presented in 165 reports, consisting of 73,498 children, were included in the meta-analyses reported here.

Coding of Studies

For all studies, we coded sample size, internal consistencies of relevant scales (for correction of unreliability), whether continuous variables were artificially dichotomized (for correction of dichotomization), and effect sizes of interest (see below). We also coded nine variables as potential moderators of effect sizes. The first of these was the method of assessing direct and indirect aggression coded as self-reports, peer nominations, peer ratings, teacher reports, parent reports, or observations. We also coded several characteristics of the sample, including the mean age of the sample, the percentage of ethnic minorities in the sample, and the country from which the sample was drawn (ultimately dichotomized into studies conducted within the United States vs. those conducted outside of the United States due to the low representation of individual countries besides the United States). When results were reported separately for boys and girls (for intercorrelations and associations with maladjustment), we coded these results both overall and by gender (to evaluate gender moderation). We also coded the gender of the first author of the study and whether the study included gender (or related terms such as *sex* or terms such as *boys* and *girls*) in the title, as these variables might be related to potential gender differences in results. To evaluate potential publication bias (described below), we coded two aspects of publication status: (a) whether the study was unpublished versus published and (b) a continuous variable of a potential hierarchy of publication quality: 1 = *unpublished, not dissertation*; 2 = *unpublished dissertation*; 3 = *published book chapter*; 4 = *published peer reviewed journal article*; and 5 = *published article in top-tier journal*, indexed as a top 20 developmental journal in impact factor according to Journal Citation Reports (e.g., *Child Development*, *Development and Psychopathology*, *Developmental Psychology*, *Journal of Abnormal Child Psychology*, *Journal of Clinical Child and Adolescent Psychology*, *Journal of Educational Psychology*, and *Journal of Experimental Child Psychology*). To assess accuracy of coding, two authors independently coded a sample of 77 studies; comparisons indicated more than 98% agreement.

We calculated effect sizes as correlations (a) of gender with each form of aggression (to evaluate gender differences), (b) between direct and indirect aggression (to evaluate intercorrelations between the two forms), or (c) of both forms of aggression with indices of psychosocial adjustment. We examined four types of adjustment indices. *Internalizing problems* included clinical or subclinical depression and anxiety. Externalizing problems included two aspects: (a) *emotional dysregulation and attention deficit hyperactivity disorder (ADHD) symptoms* such as impulsivity, frustration intolerance, tendencies to easily become angry or upset, inattention, and hyperactivity, and (b) *delinquent behaviors* such as oppositional or defiant behaviors, destruction of property, deceitfulness and theft, and rule violations. We considered these specific aspects of externalizing problems rather than a general externalizing construct to avoid overlap in behaviors that define aggression and the externalizing problem. *Prosocial behavior* included behaviors such as helping others, sharing, and cooperating. Finally, sociometric status included two aspects: *peer acceptance* (i.e., being liked by many peers) and *peer rejection* (i.e., being disliked by many peers).

Statistical Analysis

Effect size calculations. We represented most effect sizes (gender differences in direct and indirect aggression, intercorrelations between the two forms, and associations of each form of aggression with maladjustment) as Pearson correlations, r. For studies reporting results in other metrics (e.g., other effect sizes such as Cohen's d, results of significance tests such a t tests, and descriptive data such as means and standard deviations), these data were transformed to r using standard procedures (e.g., Rosenthal, 1991). Studies reporting only a significant association (at a certain p or assumed to be .05 if not otherwise stated) were given the minimum r that would achieve that level of significance given the sample size, and studies reporting only that a particular effect size was not significant (with no other information from which to compute an effect size) were assigned $r = 0$. These standard practices represent a conservative approach and may lead to slight underestimation of overall effect sizes.

In order to obtain accurate effect sizes among the constructs of interest, we corrected for two study artifacts that attenuate effect sizes, using the general equation $r_{\text{corrected}} = r/a$ (where a is the total artifactual correction, computed as the product of the following individual corrections; see Hunter & Schmidt, 2004). The first of these corrections was for unreliability of measures, in which we estimated artifactual corrections based on reported reliabilities of both variables composing the correlation, $a_{\text{reliability}} = \sqrt{\alpha_X\alpha_Y}$ (with α_x and α_Y representing the internal consistencies of variables X and Y in the correlation; Hunter & Schmidt, 2004, p. 35). For studies that did not report internal consistencies of measures, we used the mean reliability obtained from meta-analysis (Rodriguez & Maeda, 2006) of studies reporting internal consistencies ($\bar{\alpha}$s = .85 and .82 for direct and indirect aggression, respectively; full results of these analyses are available from the first author). The second correction we made was for attenuation among studies that artificially dichotomized variables used in the effect size (e.g., median or 1 *SD* splits to group participants as aggressive vs. nonaggressive). We used the artifactual correction for this dichotomization ($a_{\text{dichotomization}} = \Phi(c)/\sqrt{PQ}$; for details, see Hunter & Schmidt, 1990, 2004, p. 36) for one or both dichotomized variables comprising the correlation.

To examine the unique relations of each form of aggression to maladjustment, it was necessary to account for the magnitude of the correlation between direct and indirect aggression. To estimate the independent relation of each form of aggression (e.g., indirect aggression after controlling for direct aggression), we computed semipartial correlations (*sr*) from the (corrected) bivariate correlations of direct and indirect aggression with the adjustment variable and between direct and indirect aggression using the following formulas (Cohen & Cohen, 1983):

$$sr_{\text{direct}} = \frac{r_{\text{direct}} - r_{\text{indirect}} r_{\text{corr}}}{\sqrt{1 - r_{\text{corr}}^2}} \text{ and}$$
$$sr_{\text{indirect}} = \frac{r_{\text{indirect}} - r_{\text{direct}} r_{\text{corr}}}{\sqrt{1 - r_{\text{corr}}^2}}, \qquad (1)$$

where r_{direct} is the correlation of direct aggression to the adjustment variable, r_{indirect} is the correlation of indirect aggression to the adjustment variable, and r_{corr} is the correlation between direct and indirect aggression.

To compare the relative magnitudes of the correlations of direct versus indirect aggression with the adjustment variables, we computed a *d* score for each study representing the difference in associations (*r*s) of direct and indirect aggression to the correlate. Adapting traditional methods of comparing dependent correlation coefficients (e.g., Cohen & Cohen, 1983), the following formula was used (Card & Little, 2006):

$$d = \frac{(r_{\text{direct}} - r_{\text{indirect}})\sqrt{(n-1)(1+r_{\text{corr}})}}{\sqrt{\frac{2(n-1)}{n-1}|R| + \bar{r}(1-r_{\text{corr}})^3}} \frac{1}{\sqrt{n-3}}, \qquad (2)$$

where r_{direct} is the correlation of direct aggression to the correlate, r_{indirect} is the correlation of indirect aggression to the correlate; r_{corr} is the correlation between direct and indirect aggression, $|R| = 1 - r_{\text{direct}}^2 - r_{\text{indirect}}^2 - r_{\text{corr}}^2 + 2r_{\text{direct}} r_{\text{indirect}} r_{\text{corr}}$, $\bar{r} = (r_{\text{direct}} + r_{\text{indirect}})/2$; and n is the number of participants in study.

The semipartial correlations (*sr*) and the difference scores (*d*) both index the unique relations of direct and indirect aggression with adjustment but in different ways. The semipartial correlations indicate the direction and magnitude of the relations of each form of aggression to adjustment that are independent of the other form (specifically, the positive or negative square root of the variance in the adjustment variable that overlaps with that form of aggression only). These semipartial correlations, however, do not directly evaluate the differential relations direct and indirect forms of aggression have with the adjustment variable. This differential relation is instead captured by the difference score (*d*), which indexes the magnitude of the *difference* in the relations of each form of aggression with adjustment (but says nothing of the magnitude of association of either one of the forms of aggression with adjustment). Therefore, both the semipartial correlations and the difference scores are needed to evaluate the distinct associations direct and indirect aggression have with psychosocial adjustment.

Among studies reporting results separately for boys and girls, we were interested in whether these results differed by gender. To evaluate this potential gender moderation, we adapted traditional methods of comparing independent correlations (e.g., Cohen & Cohen, 1983) to yield the following equation:

$$d = \frac{2(Zr_{\text{boys}} - Zr_{\text{girls}})}{\left(\sqrt{\frac{1}{n_{\text{boys}}-3} + \frac{1}{n_{\text{girls}}-3}}\right)\left(n_{\text{boys}} + n_{\text{girls}} - 4\right)}, \qquad (3)$$

where Zr_{boys} is the Fisher's Z transformation of correlation for boys, Zr_{girls} is the Fisher's Z transformation of correlation for girls, n_{boys} is the number of boys in study, and n_{girls} is the number of girls in study.

This formula was used to compare the magnitude of intercorrelation between direct and indirect aggression, as well as the associations of both forms of aggression with adjustment, across gender for those studies reporting these results separately for boys and girls. In order to evaluate whether the direction and magnitude of differential associations of direct and indirect aggression with adjustment differed (i.e., was moderated) by gender, *d* scores from Equation 2 were transformed to *Zr* (for boys and girls separately), and then applied within Equation 3.

Combining and comparing effects across studies. All effect sizes were combined using weighted random-effects analyses (e.g., Hedges & Vevea, 1998). This approach was preferred because, as will be seen below, all the meta-analyses performed indicated significant heterogeneity around the mean effect sizes (i.e., fixed-effects models were not appropriate).

Before combining effect sizes, we first transformed correlation metrics (*r* and *sr*) to Fisher's Z_r (or Z_{sr}) in order to provide an approximately normally distributed metric (averaged values of Z_r and Z_{sr} were transformed back into $\bar{r}$ and $\overline{sr}$ for reporting; Rosenthal, 1991). We analyzed difference scores in their *d* metric. We averaged multiple effect sizes (Z_rs, Z_{sr}s, or *d*s) from the same study in order to yield one effect size per study (for moderator analyses, one effect size for any given level of the moderator) in order to avoid violating independence assumptions when testing significance and computing standard errors.

When averaging across multiple studies, we weighted effect sizes by an inverse variance weight using the general formula, mean ES = $\sum w\mathrm{ES}/\sum w$. Three components comprised this weight. The first component is a function of study sample size, where $w = n - 3$ for bivariate correlations; $w = n - 4$ for semipartial correlations (Hays, 1994); and $w = 1/[(4/df)(1 + (d^2/8))]$ for the differential index (Lipsey & Wilson, 2001, p. 49). The second component of this weight is an adjustment for artifact correction for the corrected effect sizes (as described by Lipsey & Wilson, 2001, pp. 110, 112). The third component of this weight is the estimated random between-study variance, τ^2 (Hedges & Vevea, 1998, pp. 491–492).

When sufficient studies existed and the overall effect sizes contained significant heterogeneity (as indexed by Q, which is distributed as χ^2 with $df = k - 1$; Lipsey & Wilson, 2001, p. 115), we systematically examined several potential moderators, including the methodology (method by which aggression was assessed), sample (sample age, gender, percentage of ethnic minorities, and country), author gender, and whether gender or similar terms appeared in the title. We also considered two publication status (dichotomous and continuous) variables as moderators in order to evaluate potential publication bias.

We evaluated categorical moderators (aggression reporter, country, first author gender, gender in title, and dichotomous publication status) using variance partitioning procedures analogous to analysis of variance (Lipsey & Wilson, 2001, pp. 135–138; Overton, 1998). Specifically, we divided heterogeneity of effect sizes (Z_rs, Z_{sr}s, or ds) among studies (Q_{total}) into within-moderator-groups (Q_{within}; e.g., among studies using teacher reports) and between-group ($Q_{\mathrm{between}} = Q_{\mathrm{total}} - \sum Q_{\mathrm{within}}$) components. This between-group heterogeneity is distributed as χ^2 (with df = number of groups − 1) under the null hypothesis that population effect sizes are equal across moderator groups. Significant between-group heterogeneity is indicative of moderation.

To test continuous moderators (i.e., sample age, percentage of ethnic minorities, and publication quality), we used procedures analogous to weighted regression (for details, see Lipsey & Wilson, 2001, pp. 138–140; Overton, 1998). Specifically, we regressed the Z_rs, Z_{sr}s, or ds from each study onto the continuous moderator, weighted by the appropriate weights described above (note that the standard errors of this weighted regression were modified according to the procedures recommended by Lipsey & Wilson, 2001, p. 140). The heterogeneity among studies accounted for by the continuous moderator, $Q_{\mathrm{regression}}$, is also distributed as χ^2 (with 1 df) under the null hypothesis that population effect sizes do not change linearly across levels of the moderator (e.g., across age). We interpreted significant relations between the moderators and the effect sizes using the effect size intercept and slope to obtain model-implied effect sizes at selected levels of the moderator (model-implied Z_rs were converted to implied rs for reporting).

As with our analyses of overall effect sizes, we modeled residual heterogeneity of effect sizes in our moderator analyses. Therefore, these analyses can be considered mixed-effects models in that the moderator variables were considered fixed, whereas effect sizes of individual studies were allowed to randomly vary (Hedges & Vevea, 1998; Overton, 1998).

Results

Before describing our results, we present general characteristics of the 148 studies included in these meta-analyses (Table 1). The studies varied considerably in terms of sample size and to a lesser extent age (with early childhood and older adolescence not well represented) and percentage of ethnic minorities in the sample (few studies had high percentages of minorities). Self-reports, peer nominations, and teacher reports were the primary methods of assessment; most studies were conducted in the United States; and most studies were published with females as first authors. Table 1 also indicates heterogeneity among studies in terms of whether the study was published (dichotomous) and publication quality.

We present the results of our meta-analyses in three parts. First, we report the magnitudes of gender differences in direct and indirect forms of aggression as well as potential moderators of these differences. Second, we investigate the intercorrelations between direct and indirect forms of aggression (and potential moderators of these intercorrelations). Third, we examine the associations of direct and indirect aggression with four indices of adjustment: internalizing problems, externalizing problems, prosocial behavior, and peer relations.

Gender Differences in Direct and Indirect Aggression

As displayed in Figure 1 (and listed in Table A1), 107 studies (consisting of 50,977 participants) reported results speaking to the magnitude of gender differences in direct and indirect aggression during childhood and adolescence. The additional studies included here but not in Archer (2004) were

Table 1

Summary of Characteristics of Studies Included in Meta-Analyses

				Percentile range[a]			
Continuous descriptors	*M*	*Mdn*	*SD*	10	25	75	90
Sample size (*N*)	574	254	1,360	74	121	526	926
Mean sample age (years)	9.8	10.0	3.34	4.4	7.6	12.0	14.0
% ethnic minority	32	28.5	23	7	15	45	66
Categorical descriptors	% Studies						
Reporter type[b]							
Self	26.4						
Peer nomination	48.6						
Peer rating	9.7						
Teacher	26.4						
Parent	7.6						
Observation	9.7						
Other	4.9						
Country[b]							
Australia	2.1						
Canada	11.8						
England	2.8						
Finland	5.6						
Germany	2.1						
United States	70.1						
Other (in less than three studies)	5.6						
Female first author	78.2						
Published	63.2						
Publication quality							
Unpublished (not dissertation)	4.2						
Unpublished dissertation	32.6						
Published book chapter	0.7						
Published journal article	43.1						
Published in top-tier journal	19.4						

[a] Values falling at 10th, 25th, 75th, and 90th percentiles. [b] Percentages do not necessarily sum to 100 because some studies included multiple categories.

predominantly published since 2000 or 2001, when the last searches for Archer's meta-analysis were performed. Table 2 summarizes the results of meta-analytic syntheses of these gender differences.

Overall, gender differences in direct aggression are generally consistent with prior reviews (e.g., Hyde, 1984; Maccoby & Jacklin, 1974). As shown in Table 2, we find significant heterogeneity among these studies in the magnitude of gender differences in direct aggression. Therefore, we conducted random-effects analyses to evaluate the typical magnitude of this gender difference and mixed-effects models when evaluating moderators. Overall, boys were more directly aggressive than girls, with an average effect that was medium in magnitude, $\bar{r} = .29$ (note that positive *r*s indicate higher levels of aggression for boys, whereas negative *r*s indicate higher levels of aggression for girls). To facilitate comparison to other meta-analyses of gender differences, this effect size is equivalent to $\bar{d} = .61$. Moderators of these effect sizes revealed several noteworthy findings. First, method of assessing direct aggression is related to the magnitude of gender differences, with parent reports and self-reports yielding the smallest gender differences, whereas peer reports (nominations or ratings) and observations yield the largest differences (see Table 2 for full details). The percentage of ethnic minorities in samples also relates to the magnitude of gender differences in direct aggression, such that smaller differences are found with increasing percentages of ethnic minorities. Furthermore, studies with the word *gender* (or similar terms) in the title yield slightly larger gender differences in direct aggression, although studies without this term in their title still exhibit significant, moderate magnitude differences. The mean age of the sample, country from which the

Gender Differences (r)		
Direct Aggression		Indirect Aggression
8	.9	
	.8	
	.7	
20	.6	
632100	.5	4
887663322111000	.4	0
99877776666544321110000000	.3	245
9988887766665554333221000	.2	066
99988777776643333322 11	.1	0001112345
998766430	.0	00111112222334456888888
10	-.0	001133335566666666788889
	-.1	0001123333334445778
	-.2	0011122345556688
	-.3	225
	-.4	138
	-.5	
.29	$\bar{r}$	-.03
869.2	$Q_{(106)}$	574.3

Figure 1. Stem-and-leaf plot of gender differences in direct and indirect aggression.

Table 2

Summary of Meta-Analytic Results of Gender Differences in Direct and Indirect Aggression

	Direct	Indirect
Heterogeneity (Q; $df = 106$)	869.2***	574.3***
Random-effects mean r	.29***	−.03*
95% CI	0.26:0.32	−0.05:−0.01
Equivalent d	.61	−.06
95% CI	0.55:0.67	−0.11:−0.02
Moderators		
Reporter ($Q_{b(5)}$)	80.17***	28.66***
Observation ($k = 14$)	.37***$_{\text{cd}}$	−.05$_{\text{ab}}$
Parent ($k = 8$)	.15***$_{\text{a}}$	−.08***$_{\text{a}}$
Peer nomination ($k = 39$)	.38***$_{\text{d}}$	−.02$_{\text{ab}}$
Peer ratings ($k = 12$)	.37***$_{\text{d}}$	.01$_{\text{ab}}$
Self-report ($k = 33$)	.21***$_{\text{b}}$	.03*$_{\text{b}}$
Teacher ($k = 31$)	.34***$_{\text{bc}}$	−.07**$_{\text{a}}$
Age ($Q_{\text{regression}(1)}$)	0.01	0.22
β_{Zr}	−.001	.002
$\hat{r}$ at 5, 10, and 15 years	.29:.29:.29	−.04:−.04:−.03
% ethnic minority ($Q_{\text{regression}(1)}$)	4.70*	0.22
β_{Zr}	−.0016	.0003
$\hat{r}$ at 25%, 50%, and 75%	.30:.26:.22	−.05:−.04:−.03
Country ($Q_{b(1)}$)^{a}	0.01	2.72
United States ($k = 72$)	.29***	−.01
Other countries ($k = 34$)	.29***	−.05**
First author gender ($Q_{b(1)}$)	3.59	2.48
Female ($k = 83$)	.31***	−.06***
Male ($k = 22$)	.25***	−.01
Gender in title ($Q_{b(1)}$)^{b}	4.92*	0.09
No ($k = 72$)	.27***	−.04*
Yes ($k = 35$)	.34***	−.03

Note. Significant differences among reporters from follow-up comparisons are denoted by different alphabetic subscripts. CI = confidence interval.

[a] Due to low representation of individual countries outside of the United States, this variable was dichotomized to studies conducted within versus outside of the United States. [b] Dichotomous variable of whether the word *gender* (or similar terms such as *sex* or *boys* and *girls*) was in the title.

*$p < .05$. **$p < .01$. ***$p < .001$.

sample was drawn, and first author gender are not significantly associated with magnitudes of gender differences in direct aggression. In order to compare gender differences in physical versus verbal aspects of direct aggression, we conducted post hoc analyses of 27 studies that reported gender differences separately for these two aspects. Results indicate that gender differences are stronger for physical ($\bar{r} = .34$, equivalent $\bar{d} = .73$) than verbal ($\bar{r} = .19$, equivalent $\bar{d} = .38$) expressions of direct aggression.

We also find significant heterogeneity across studies for gender differences in indirect aggression. Similar to Archer's (2004) findings, our (random effects) results indicate a negligible, but statistically different from zero, average gender difference, with girls exhibiting more indirect aggression than boys, $\bar{r} = -.03$ (equivalent $\bar{d} = -.06$). As with direct aggression, this gender difference in indirect aggression varies by reporter. Specifically, parent and teacher reports yield gender differences of girls being higher than boys, whereas self-reports yield a slightly higher level for boys than girls. Other reports are not significantly different from zero (see Table 2 for details). However, the magnitude of these gender differences in indirect aggression is trivial regardless of reporter. Age, percentage of ethnic minorities in sample, country of sample, first author gender, and whether gender appeared in the title do not moderate gender differences in indirect aggression.

Intercorrelations Between Direct and Indirect Aggression

Figure 2 and Table A2 show the 98 studies reporting associations between direct and indirect aggression. As shown in Table 3, we find significant heterogeneity among these effect sizes. The random-effects average corrected correlation between the two forms of aggression is $\bar{r}$ = .76 (even uncorrected effect sizes representing correlations as found in studies is high, $\bar{r}$ = .60; results involving uncorrected effect sizes mirror those presented here and are available from the first author). Although separable (i.e., the confidence interval of this correlation does not include 1.0), direct and indirect aggression are clearly strongly correlated.

Given the heterogeneity of effect sizes across studies, we examined moderators of this association (Table 3). Method of assessment significantly moderates this association, such that observational methods yield considerably lower correlations than other methods. Direct and indirect aggression are also more strongly correlated among boys than girls, as well as with increasing percentages of ethnic minorities in the samples. The mean age of the sample, country from which the sample was drawn, the gender of the first author of the study, and whether the source included the term *gender* in the title do not significantly moderate the magnitude of correlations between direct and indirect aggression.

	Direct-Indirect *r*
1.0	00
.9	122244556779
.8	00033445567899
.7	0001223334455666677789
.6	01222223345669999999
.5	0356888
.4	02556678
.3	1169
.2	4
.1	1149
.0	78
-.0	
$\bar{r}$	.76
$Q_{(72)}$	5416.13

Figure 2. Stem-and-leaf plot of intercorrelations between direct and indirect aggression.

Associations of Direct and Indirect Aggression With Maladjustment

We next consider the magnitudes of associations of direct and indirect aggression with four aspects of adjustment: internalizing problems, externalizing problems, prosocial behavior, and peer relations. We first present average bivariate correlations of each form of aggression with the adjustment index, then average independent (i.e., semipartial) correlations for each form and the differential association (i.e., *d*) that direct versus indirect aggression has with the adjustment index. As before, we systematically evaluate potential moderation of these effects.

Table 3

Summary of Meta-Analytic Results of Correlations Between Direct and Indirect Aggression

Heterogeneity (Q; df = 97)	5,416.13***
Random-effects mean *r*	.76***
95% CI	0.72:0.79
Moderators	
Reporter ($Q_{b(5)}$)	61.93***
Observation (k = 9)	.33***$_{a}$
Parent (k = 7)	.74***$_{b}$
Peer nomination (k = 49)	.84***$_{c}$
Peer ratings (k = 7)	.77***$_{bc}$
Self-report (k = 22)	.73***$_{b}$
Teacher (k = 28)	.66***$_{b}$
Age ($Q_{\text{regression}(1)}$)	1.36
β_{Zr}	.015
$\hat{r}$ at 5, 10, and 15 years	.73:.76:.79
Gender (k = 43) (d)^{a}	.21***
Girls	.82***
Boys	.88***
% ethnic minority ($Q_{\text{regression}(1)}$)	5.24*
β_{Zr}	.0046
$\hat{r}$ at 25%, 50%, and 75%	.73:.78:.82
Country ($Q_{b(1)}$)^{b}	0.00
United States (k = 72)	.76***
Other countries (k = 26)	.76***
First author gender ($Q_{b(1)}$)	0.79
Female (k = 77)	.74***
Male (k = 21)	.79***
Gender in title ($Q_{b(1)}$)^{c}	1.31
No (k = 69)	.74***
Yes (k = 29)	.80***

Note. Significant differences among reporters from follow-up comparisons are denoted by different alphabetic subscripts. CI = confidence interval.

[a] See Equation 3 and accompanying text for description of how gender moderation was evaluated. [b] Due to low representation of individual countries outside of the United States, this variable was dichotomized to studies conducted within versus outside of the United States. [c] Dichotomous variable of whether the word *gender* (or similar terms such as *sex* or *boys* and *girls*) was in title.

*p < .05. ***p < .001.

Internalizing Problems (r)		
Direct Aggression		Indirect Aggression
	.5	
	.4	125
984	.3	347
950	.2	
991	.1	00038
9766540	.0	00245579
772110	-.0	0223
30	-.1	9
1	-.2	4
	-.3	5
0	-.4	
.07	$\bar{r}$	.10
468.47	$Q_{(25)}$	635.61

Figure 3. Stem-and-leaf plot of associations of direct and indirect aggression with internalizing problems.

Internalizing problems. Results of 26 studies (Figure 3 and Table A3) indicate significant heterogeneity in the bivariate, semipartial, and differential associations of direct and indirect aggression with internalizing problems (Table 4). Only indirect aggression exhibits significant bivariate and unique (i.e., semipartial) associations with internalizing problems as well as significant differential association between the two forms with internalizing problems. Moderator analyses reveal three differences. The first difference is by reporter, such that parent, teacher, and self-reported direct and indirect aggression exhibit the strongest associations with internalizing problems, although aggression reporter does not significantly moderate the differential association of direct versus indirect aggression with internalizing problems. The second significant moderator is the percentage of ethnic minorities in the sample, such that samples with higher percentages of ethnic minorities exhibit higher differential associations between direct versus indirect aggression. This effect is explained by nonsignificant trends of indirect aggression having stronger positive associations and direct aggression having stronger negative associations with higher percentages of minorities. The third significant moderator is whether the title of the work included *gender* or similar terms. Studies with *gender* in the title show stronger differential associations of the two forms of aggression with internalizing problems; specifically, studies with this term have positive unique (semipartial) associations of indirect aggression to internalizing problems and unique links of direct aggression to *lower* internalizing problems. Sample age, participant gender, country in which the study was conducted, and first author gender (with one exception) do not consistently moderate associations of direct or indirect aggression with internalizing problems.

Externalizing problems. As mentioned, we report results regarding two aspects of externalizing problems: emotional dysregulation/ADHD-type symptoms and delinquency/conduct problems. This focus on specific aspects of externalizing problems allows us to avoid a potential confound that would exist if we had considered general externalizing problems, which are often measured using items that also assess aggression. Of the 18 studies considering emotional dysregulation and ADHD-type symptoms, 12 specifically assessed symptoms of ADHD, 5 assessed ease of anger or frustration, and 1 assessed a more general emotional reactivity. Of the 14 studies of delinquency/conduct problems, 10 used established measures of conduct problems (e.g., Child Behavior Checklist conduct problems subscale) and 4 used specific behavioral descriptions. The scales used to assess each type of externalizing problems were generally well-established measures designed to specifically measure these constructs. For instance, the measures from Achenbach's system (i.e., Child Behavior Checklist, Youth Self Report, and Teacher Report Form), used in several of these studies, explicitly distinguish aggression from both delinquent behaviors (McMahon & Estes, 1997) and ADHD symptoms (Barkley, 1997). For these reasons, it is reasonable to consider associations of direct and indirect aggression with these conceptually distinct aspects of externalizing problems.

The 18 studies investigating associations with emotional dysregulation and ADHD-type symptoms (see left portion of Figure 4 and upper portion of Table A4) exhibit significant heterogeneity in their results (Table 5). Random-effects combination of these results indicates that both direct and indirect forms of aggression have large bivariate correlations to these symptoms. However, direct aggression is more strongly associated than indirect aggression (differential *d* in Table 5), and only direct aggression is uniquely related to these symptoms (semipartial correlations in Table 5). The source of aggression measurement moderates these associations, such that (a) parent-reported direct aggression is most strongly linked to these symptoms and (b) both parent and teacher reports exhibit the largest differential associations with direct being more strongly associated than indirect. We also find that studies with the term *gender* in the title exhibit greater differential

Table 4

Summary of Meta-Analytic Results of Associations of Direct and Indirect Aggression With Internalizing Problems

	Direct		Indirect		
	r	*sr*	*r*	*sr*	Differential *d*
Heterogeneity (Q; $df = 25$)	468.47***	165.80***	635.61***	316.60***	76.86***
Random-effects mean ES	.07	−.01	.10*	.08*	−.11**
95% CI	−0.02:0.15	−0.07:0.04	0.01:0.19	0.02:0.14	−0.19:−0.03
Moderators					
Reporter ($Q_{b(5)}$)	77.75***	29.42***	89.14***	38.40***	6.61
Observation ($k = 1$)	−.10$_{a}$	−.16$_{a}$	.00$_{a}$	.12$_{ab}$	−.30
Parent ($k = 1$)	.32***$_{c}$	.12***$_{b}$	.40***$_{c}$	.26***$_{c}$	−.18***
Peer nomination ($k = 8$)	−.06$_{a}$	−.06$_{a}$	−.04$_{a}$	.01$_{a}$	−.07
Peer ratings ($k = 1$)	.00$_{a}$	.00$_{a}$	.00$_{a}$	.00$_{a}$	.00
Self-report ($k = 7$)	.12*$_{b}$	−.02$_{a}$	.17*$_{b}$	.13*$_{b}$	−.16*
Teacher ($k = 8$)	.20**$_{b}$	.03$_{a}$	.23***$_{b}$	.14**$_{b}$	−.08
Age ($Q_{\text{regression}(1)}$)	0.79	1.03	0.28	0.08	0.14
β_{ES}	−.012	−.009	−.008	−.003	−.005
ES at 5, 10, and 15 years	.12:.06:.00	.03:−.02:−.06	.14:.10:.06	.09:.08:.06	−.08:−.11:−.13
Gender ($k = 10$) (d)[a]	.03	.07	.00	−.04	.00
Girls	−.02	−.05	.05	.07	−.01
Boys	.05	.04	.06	−.01	−.01
% ethnic minority ($Q_{\text{regression}(1)}$)	1.38	0.58	3.19	3.75	6.10*
β_{ES}	.0019	−.0007	.0032	.0022	−.0032
ES at 25%, 50%, and 75%	.05:.09:.14	−.02:−.03:−.05	.07:.15:.22	.06:.11:.17	−.09:−.16:−.24
Country ($Q_{b(1)}$)[b]	0.00	0.86	0.09	0.37	0.20
United States ($k = 22$)	.07	−.02	.10*	.09*	−.12**
Other countries ($k = 5$)	.07	.02	.06	.03	−.02
First author gender ($Q_{b(1)}$)	3.08	0.00	3.92*	3.55	2.01
Female ($k = 24$)	.08	−.01	.11*	.08*	−.11**
Male ($k = 3$)	−.02	−.01	−.01	.00	−.02
Gender in title ($Q_{b(1)}$)[c]	0.01	4.99*	2.15	5.97*	14.82***
No ($k = 17$)	.07	.02	.06	.02	−.01
Yes ($k = 10$)	.07	−.09*	.19**	.19***	−.31***

Note. Significant differences among reporters from follow-up comparisons are denoted by different alphabetic subscripts. ES = effect size; CI = confidence interval.

[a] See Equation 3 and accompanying text for description of how gender moderation was evaluated. [b] Due to low representation of individual countries outside of the United States, this variable was dichotomized to studies conducted within versus outside of the United States. [c] Dichotomous variable of whether the word *gender* (or similar terms such as *sex* or *boys* and *girls*) was in title.

*$p < .05$. ** $p < .01$. *** $p < .001$.

associations (with direct being more strongly associated than indirect) than studies without this term. However, the source of this moderation (i.e., studies with this term have stronger unique associations for direct aggression, whereas studies without this term had stronger unique associations for indirect aggression) is not readily explainable. Participant age and gender, percentage of ethnic minorities in sample, and author gender do not significantly moderate any of the associations (bivariate, semipartial, or differential) of these forms of aggression with emotional dysregulation/ADHD-type symptoms.

The results of 14 studies reporting associations of direct and indirect aggression with delinquency and conduct problems (shown in the right of Figure 4 and bottom of Table A4) are also significantly heterogeneous (Table 6). Random-effects meta-analyses indicate that both direct and indirect aggression have large bivariate correlations with these problems and that both are uniquely associated with these problems. The differential index indicates that direct is more strongly related to these problems than indirect aggression, as evidenced by the medium unique effect of direct in contrast to the small unique effect of indirect. These results are moderated by method of assessing aggression (Table 6). However, these reporter moderator effects are due primarily to the single study using parent-reported direct and indirect aggression

Emotional Dysregulation / ADHD Symptoms			Delinquency / Conduct Disorder		
Direct Aggression		Indirect Aggression	Direct Aggression		Indirect Aggression
0	.9		20	.9	
	.8	15		.8	
	.7			.7	019
955540	.6	3	9622	.6	2
21	.5	23	9.	5	07
60	.4	235	92	.4	12
84	.3	489	60	.3	45
84	.2	03	71	.2	0
	.1	3	8.	1	5
70	.0	08		.0	7
	-.0	0		-.0	4
	-.1			-.1	
.52	$\bar{r}$	.42	.58	$\bar{r}$	.45
505.14	$Q_{(17)}$	550.82	850.41	$Q_{(13)}$	371.40

Figure 4. Stem-and-leaf plot of associations of direct and indirect aggression with externalizing problems.

(Tiet, Wasserman, Loeber, McReynolds, & Miller, 2001). If this study is removed, there is no evidence of reporter moderation. The only other moderator of these results is the percentage of ethnic minorities in samples. Here, indirect aggression exhibits increasingly stronger (bivariate and semipartial) associations with conduct problems in samples with increasing percentages of ethnic minorities (this effect also results in smaller differentials between direct and indirect aggression in samples with higher percentages of ethnic minorities). Participant age and gender, country of sample origin, author gender, and whether *gender* appeared in the title do not moderate these results.

Prosocial behavior. Associations of direct and indirect aggression with prosocial behavior are significantly heterogeneous across the 31 studies reporting these effects (Table 7, Figure 5, and Table A5). Random-effects meta-analyses indicate that both direct and indirect aggression have negative bivariate correlations with prosocial behavior (Table 7). However, the differential association (*d*) is significant, and this difference can be seen clearly in the unique (semipartial) associations: Direct aggression is uniquely related to *low* prosocial behavior, whereas indirect aggression is uniquely related to *high* prosocial behavior.

Several features moderate these associations of direct and indirect aggression with prosocial behavior (Table 7). Age moderates the differential between direct and indirect aggression, such that the strongest differences (of direct aggression being associated with low prosocial behavior, whereas indirect aggression is associated with high prosocial behavior) are evident at younger ages. Method of assessing aggression consistently moderates these associations (Table 7). The country from which the sample was drawn also moderates effects, such that indirect aggression only exhibits unique (semipartial) associations to *high* prosocial behavior in the United States; samples drawn from other countries yield a negative semipartial correlation between indirect aggression and prosocial behavior.

Peer relations. Two aspects of peer relations have been examined with respect to direct and indirect aggression in a sufficient number of studies to allow meta-analytic examination: peer acceptance and peer rejection.

The 23 studies (see left portion of Figure 6 and top of Table A6) reporting associations of direct and indirect aggression with peer acceptance are significantly heterogeneous (Table 8). Both forms of aggression have small to moderate bivariate associations with low peer acceptance, but this association differs by form, and only direct aggression is uniquely associated with low peer acceptance. Moderator analyses yield only two findings. First, indirect aggression more strongly relates to low peer acceptance with increasing percentages of ethnic minorities in samples (this trend also results in changes in the direct vs.

Table 5

Summary of Meta-Analytic Results of Associations of Direct and Indirect Aggression With Emotional Dysregulation and ADHD Symptoms

	Direct		Indirect		
	r	*sr*	*r*	*sr*	Differential *d*
Heterogeneity (Q; $df = 17$)	505.14***	435.01***	550.82***	419.32***	213.53***
Random-effects mean ES	.52***	.31***	.42***	.06	.35***
95% CI	0.42:0.61	0.20:0.41	0.30:0.53	−0.06:0.17	0.17:0.52
Moderators					
Reporter ($Q_{b(4)}$)	18.16**	8.56	10.17*	6.28	108.44***
Observation ($k = 1$)	.00$_a$	.00	.00$_a$	.00	.00$_a$
Parent ($k = 3$)	.89***$_c$	.82**	.46**$_b$	−.22	1.15***$_c$
Peer nomination ($k = 3$)	.41***$_b$	.19	.50*$_b$	.23	−.17$_a$
Peer ratings ($k = 0$)	—	—	—	—	—
Self-report ($k = 5$)	.51***$_b$	.21**	.51***$_b$	.13*	.08$_a$
Teacher ($k = 8$)	.53***$_b$	.41***	.33***$_b$	−.03	.66***$_b$
Age ($Q_{\text{regression}(1)}$)	2.64	0.28	2.17	0.07	0.40
β_{ES}	.034	.010	.032	−.005	.017
ES at 5, 10, and 15 years	.43:.56:.66	.28:.32:.36	.33:.46:.58	.07:.05:.03	.29:.37:.46
Gender ($k = 7$) (d)^a	.03	.19	.02	−.11	.03
Girls	.42***	.13	.32***	.08	.01
Boys	.48***	.31**	.38***	−.02	.05
% ethnic minority ($Q_{\text{regression}(1)}$)	1.65	0.50	0.19	0.67	2.59
β_{ES}	.0035	.0017	.0012	−.0019	.0055
ES at 25%, 50%, and 75%	.51:.57:.63	.33:.36:.40	.39:.41:.44	.03:−.02:−.06	.42:.58:.69
Country ($Q_{b(1)}$)^b	0.01	0.27	0.01	0.05	0.51
United States ($k = 14$)	.55***	.33***	.44***	.07	.37***
Other countries ($k = 3$)	.53***	.28***	.43**	.05	.26***
First author gender ($Q_{b(1)}$)	1.86	2.28	0.98	0.17	0.87
Female ($k = 12$)	.57***	.36***	.46***	.08	.41**
Male ($k = 6$)	.41***	.20**	.37**	.04	.24*
Gender in title ($Q_{b(1)}$)^c	0.61	2.01	0.62	4.04*	14.87***
No ($k = 10$)	.48***	.24***	.46***	.15*	.07
Yes ($k = 8$)	.57***	.39***	.36**	−.07	.75***

Note. Significant differences among reporters from follow-up comparisons are denoted by different alphabetic subscripts. ADHD = attention deficit hyperactivity disorder; ES = effect size; CI = confidence interval.

[a] See Equation 3 and accompanying text for description of how gender moderation was evaluated. [b] Due to low representation of individual countries outside of the United States, this variable was dichotomized to studies conducted within versus outside of the United States. [c] Dichotomous variable of whether the word *gender* (or similar terms such as *sex* or *boys* and *girls*) was in title.

$^*p < .05$. $^{**}p < .01$. $^{***}p < .001$.

indirect differential with ethnic composition; Table 8). Second, we find a similar moderator effect for whether the term *gender* appears in the title, such that stronger associations of indirect aggression with low peer acceptance are found for studies with this term. Other moderators (i.e., sample age, method of assessing aggression, gender, country, and author gender) do not moderate links of direct and indirect aggression with peer acceptance.

The 22 studies (see right of Figure 6 and bottom of Table A6) reporting associations of direct and indirect aggression with peer rejection are also heterogeneous (Table 9). Both forms exhibit significant bivariate and semipartial associations with peer rejection, although this link is stronger for direct than indirect aggression. We also find several moderators of these effects. Method of assessing aggression moderates some effects (Table 9) such that parent-, teacher-, and peer-reported aggression (direct and indirect) are more strongly linked than observations of aggression. Moderation by country is also evident for the bivariate (but not semipartial or differential) associations, such that both direct and indirect aggression are more strongly related to peer rejection within the United States than in other countries. Finally, female first authors find significant unique (semipartial) associations of indirect aggression with peer rejection, whereas male first authors do not find this effect

Table 6

Summary of Meta-Analytic Results of Associations of Direct and Indirect Aggression With Delinquent Behaviors

	Direct		Indirect		
	r	sr	r	sr	Differential d
Heterogeneity (Q; df = 13)	850.41\	365.57***	371.40***	151.20***	278.31***
Random-effects mean ES	.58***	.27***	.45***	.11**	.27*
95% CI	0.43:0.69	0.15:0.38	0.34:0.55	0.03:0.19	0.06:0.49
Moderators					
Reporter ($Q_{b(3)}$)	7.82*	64.75***	29.79***	42.49***	106.13***
Observation (k = 0)	—	—	—	—	—
Parent (k = 1)	.59***$_b$	−.32**$_a$	.70***$_b$	.49***$_b$	−1.26***$_a$
Peer nomination (k = 3)	.37***$_a$	.13$_b$	.34***$_a$	.07$_a$	.07$_b$
Peer ratings (k = 0)	—	—	—	—	—
Self-report (k = 6)	.48***$_{ab}$	.27***$_b$	.40***$_a$	.10$_a$	.21**$_b$
Teacher (k = 3)	.69**$_b$	.42*$_b$	.49***$_a$	.07$_a$	.75*$_b$
Age ($Q_{\text{regression}(1)}$)	2.45	3.65	0.14	1.61	9.63**
β_{ES}	−.049	−.037	−.008	.016	−.107
ES at 5, 10, and 15 years	.74:.60:.42	.45:.30:.12	.49:.46:.43	.02:.10:.18	.90:.36:−.17
Gender (k = 4) (d)[a]	−.04	.12	−.09	−.09	.09
Girls	.47**	.08	.40**	.14	−.14
Boys	.45***	.18	.33	.07	.01
% ethnic minority ($Q_{\text{regression}(1)}$)	0.49	1.15	3.91*	4.72*	3.93*
β_{ES}	.0022	−.0021	.0042	.0028	−.0069
ES at 25%, 50%, and 75%	.56:.59:.63	.32:.28:.23	.39:.47:.55	.06:.13:.19	.44:.27:.10
Country ($Q_{b(1)}$)[b]	6.08*	1.95	2.70	0.59	1.07
United States (k = 11)	.63***	.31**	.49***	.13*	.34*
Other countries (k = 3)	.34***	.15**	.31***	.06	.11
First author gender ($Q_{b(1)}$)	0.02	0.07	0.03	0.00	0.01
Female (k = 10)	.57***	.28**	.46***	.12**	.24**
Male (k = 4)	.60**	.23*	.43**	.12	.28*
Gender in title ($Q_{b(1)}$)[c]	0.05	1.69	1.31	2.08	3.02
No (k = 10)	.56***	.34**	.40***	.06	.44**
Yes (k = 4)	.60**	.09	.56***	.23*	.16

Note. Significant differences among reporters from follow-up comparisons are denoted by different alphabetic subscripts. ES = effect size; CI = confidence interval.

[a] See Equation 3 and accompanying text for description of how gender moderation was evaluated. [b] Due to low representation of individual countries outside of the United States, this variable was dichotomized to studies conducted within versus outside of the United States. [c] Dichotomous variable of whether the word *gender* (or similar terms such as *sex* or *boys* and *girls*) was in title.

*$p < .05$. **$p < .01$. ***$p < .001$.

(resulting also in male first authors finding differential effects between direct and indirect aggression that are not found by female first authors).

Publication Bias

Publication bias is a threat to any meta-analytic review (for overviews, see Begg, 1994; Rothstein, Sutton, & Borenstein, 2005). The threat of this bias is that unpublished studies, relative to published studies, are more likely to have nonsignificant results (or otherwise smaller effects) and less likely to be included in a meta-analytic review. The result of this possibility is that the conclusions drawn from this biased sample are themselves biased, specifically with estimated effect sizes being larger than those that actually exist. The threat of publication bias can be managed in a number of ways. The most important way is to obtain unpublished studies, which we have done by including unpublished dissertations and requesting unpublished studies from researchers in the field. However, even this does not ensure that publication bias is not evident. Therefore, we statistically evaluate the potential presence of publication bias in three ways.

Differences by publication status. The first way we evaluated potential publication bias was to test whether obtained results differ by publication status. As mentioned above, we considered publication

Table 7
Summary of Meta-Analytic Results of Associations of Direct and Indirect Aggression With Prosocial Behaviors

	Direct		Indirect		
	r	*sr*	*r*	*sr*	Differential *d*
Heterogeneity (Q; $df = 30$)	422.70***	597.63***	403.90***	578.43***	246.59***
Random-effects mean ES	−.29***	−.27***	−.14***	.11**	−.42***
95% CI	−0.35:−0.22	−0.35:−0.20	−0.21:−0.08	0.03:0.18	−0.53:−0.31
Moderators					
Reporter ($Q_{b(5)}$)	17.90**	13.16*	13.93**	14.96**	23.19***
Observation ($k = 2$)	−.04$_{a}$	.00$_{a}$	−.05$_{a}$	−.07$_{ab}$	.07$_{a}$
Parent ($k = 2$)	−.36***$_{b}$	−.23***$_{bc}$	−.28**$_{b}$	−.09$_{a}$	−.16**$_{ab}$
Peer nomination ($k = 17$)	−.31***$_{b}$	−.24***$_{bc}$	−.24***$_{b}$	.07$_{bc}$	−.34***$_{b}$
Peer ratings ($k = 0$)	—	—	—	—	—
Self-report ($k = 6$)	−.10*$_{a}$	−.16**$_{b}$	.10$_{a}$	.38**$_{d}$	−.18**$_{ab}$
Teacher ($k = 12$)	−.39***$_{b}$	−.36***$_{c}$	−.10$_{a}$	.15*$_{cd}$	−.65***$_{c}$
Age ($Q_{\text{regression}(1)}$)	0.96	2.69	0.53	2.35	7.61**
β_{ES}	.011	.021	−.008	−.019	.049
ES at 5, 10, and 15 years	−.32:−.28:−.22	−.35:−.25:−.15	−.11:−.16:−.20	.18:.08:−.01	−.61:−.36:−.11
Gender ($k = 12$) (d)[a]	.04	.01	−.03	−.01	.01
Girls	−.18*	−.21*	.03	.16	−.28
Boys	−.15	−.23*	−.01	.17	−.23
% ethnic minority ($Q_{\text{regression}(1)}$)	0.28	0.67	1.19	0.78	1.02
β_{ES}	−.0008	−.0014	−.0017	.0015	−.0024
ES at 25%, 50%, and 75%	−.28:−.30:−.32	−.26:−.29:−.33	−.12:.−.16:−.20	.10:.14:.17	−.41:−.47:−.53
Country ($Q_{b(1)}$)[b]	0.70	3.03	0.95	8.64**	11.96***
United States ($k = 27$)	−.30***	−.30***	−.13***	.13*	−.51***
Other countries ($k = 4$)	−.25***	−.16**	−.19***	−.06*	−.12
First author gender ($Q_{b(1)}$)	4.97*	8.34**	0.82	0.09	9.88**
Female ($k = 26$)	−.32***	−.31***	−.16***	.11**	−.48***
Male ($k = 5$)	−.10	−.09	−.05	.08	−.14
Gender in title ($Q_{b(1)}$)[c]	0.66	1.11	13.81***	6.42*	3.74
No ($k = 21$)	−.27***	−.31***	−.07	.18***	−.47***
Yes ($k = 10$)	−.34***	−.19	−.31***	−.07	−.13

Note. Significant differences among reporters from follow-up comparisons are denoted by different alphabetic subscripts. ES = effect size; CI = confidence interval.
[a]See Equation 3 and accompanying text for description of how gender moderation was evaluated. [b]Due to low representation of individual countries outside of the United States, this variable was dichotomized to studies conducted within versus outside of the United States. [c]Dichotomous variable of whether the word *gender* (or similar terms such as *sex* or *boys* and *girls*) was in title.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

status in terms of both a dichotomous unpublished versus published variable as well as a more continuous publication quality variable. As summarized in the first column of Table 10, we compared unpublished versus published studies as moderators of each of the 33 effect sizes considered in this review. We similarly evaluated publication quality (unpublished studies other than dissertations, dissertation, chapter, journal article, and article in top-tier journal) as a continuous moderator of these effects (second column of Table 10).

Of these 66 analyses, 11 are significant. Examination of Table 10 shows that these significant moderators involve associations of direct and indirect aggression with prosocial behavior and peer rejection. For prosocial behavior, bivariate and semipartial associations of direct aggression and bivariate associations of indirect aggression are moderated by both the dichotomous and the continuous publication status variable (the latter also moderates the differential association). However, the direction of moderation in each case is such that unpublished/low-status studies exhibit *stronger* associations with low prosocial behavior (i.e., more negative) than published/high-status studies. Similar directions are evident for peer rejection, in which the bivariate correlations for both direct and indirect aggression are greater for unpublished/low-status studies than for published/high-status

Prosocial Behavior		
Direct Aggression		Indirect Aggression
	.4	
	.3	13
5	.2	349
	.1	3
110	.0	0189
610	-.0	056
75321	-.1	789
955	-.2	015579
99754	-.3	16889
8510	-.4	04
9655310	-.5	7
	-.6	0
	-.7	
-.29	$\bar{r}$	-.14
422.70	$Q_{(30)}$	403.90

Figure 5. Stem-and-leaf plot of associations of direct and indirect aggression with prosocial behavior.

studies. Given the general lack of moderation by publication status, as well as the unexpected direction of those that do emerge, these analyses fail to indicate a threat of publication bias.

Sample size–effect size correlations. A second way we tested for publication bias is by examining correlations between sample sizes and effect sizes in studies. If publication bias—in the form of nonsignificant results being less likely to be included in analyses—exists, then one would expect a correlation between these two variables (a negative correlation if the overall effect size is positive and a positive correlation if the overall effect size is negative). The reason for this correlation would be that studies with small samples would need larger effect sizes to obtain statistical significance. Therefore, only those that happened to obtain these large effects would be published (or otherwise included in the meta-analysis), whereas those that found smaller effects would be unpublished/excluded from the meta-analysis. In contrast, studies with large sample sizes would be more likely to obtain significance regardless of effect size (assuming that the population effect size itself is nonzero). Thus, a correlation between study sample size and effect size can indicate the presence of publication bias.

We computed correlations between sample sizes and effect sizes for each of the 33 effects in our review. As can be seen in Table 10 (third column), the majority of these correlations are not significant. In fact, only two related sample size/effect size correlations are significant: (a) the semipartial correlation of direct aggression with peer rejection and (b) the differential association of direct versus indirect aggression with peer rejection. Visual inspection of funnel plots (available from first author upon request) suggests that two studies (Johnson, 2003; Nelson, Robinson, & Hart, 2005) primarily contributed to these correlations.

Failsafe numbers. A third method of evaluating the potential impact of publication bias is to compute failsafe numbers (Becker, 2005; Rosenthal, 1979). Failsafe numbers indicate the number of studies with average effect sizes equal to zero that would have to exist to conclude nonsignificant effects. This number is meant to index the number of studies with effect sizes of zero that could have been excluded from the meta-analysis before the conclusions of significance would be invalidated. If this failsafe number is small, then the findings are not robust because the possibility of just a small number of excluded studies could change the conclusions of the meta-analysis. However, if the number is large, then one can be confident that the findings of the review are robust even if a large number (i.e., the failsafe number) of studies, all with an average effect size of zero, were inadvertently excluded from the meta-analysis. A common rule of thumb applied to determining whether the failsafe number is adequately large is $N_{\text{min}} = 5k + 10$ (where k is the number of studies currently in meta-analysis; Becker, 2005; Rosenthal, 1979).

We computed Rosenthal's (1979) failsafe N for all effect sizes that were statistically significant in our review (it is not relevant to compute this number for nonsignificant effects because addition of null findings would not change the nonsignificant result). We used the formula: $N = k(Z_s/Z_\alpha)^2 - k$, where k is the number of studies in the meta-analysis, Z_s is the summed significance level of studies in the meta-analysis ($Z_s = \sum Z_i/\sqrt{k}$), and Z_α is the critical Z value at $\alpha = .05$ (Becker, 2005; Rosenthal, 1979). As can be seen in Table 10 (fourth and fifth columns), the failsafe numbers exceed recommended minimums in all but one case (the failsafe number for the semipartial correlation of indirect aggression with prosocial behavior does not exceed this recommended minimum).

The results of all three approaches to evaluating publication bias converge to the general conclusion that publication bias is not a likely threat to the results of this review.

Table 8

Summary of Meta-Analytic Results of Associations of Direct and Indirect Aggression With Peer Acceptance

	Direct		Indirect		
	r	sr	r	sr	Differential d
Heterogeneity (Q; df = 22)	136.38***	160.14***	229.48***	321.81***	73.16***
Random-effects mean ES	−.17***	−.11**	−.12*	.01	−.15**
95% CI	−0.23:−0.10	−0.18:−0.04	−0.20:−0.03	−0.08:0.10	−0.25:−0.05
Moderators					
Reporter ($Q_{b(3)}$)	5.45	6.52	3.98	3.08	4.37
Observation (k = 2)	.04	.10	−.03	−.11	.24
Parent (k = 1)	−.31***	−.23***	−.22**	−.06	−.21
Peer nomination (k = 17)	−.15***	−.07	−.12*	.00	−.10
Peer ratings (k = 0)	—	—	—	—	—
Self-report (k = 0)	—	—	—	—	—
Teacher (k = 7)	−.18*	−.19**	−.06	.06	−.26***
Age ($Q_{\text{regression}(1)}$)	2.54	3.42	0.29	0.17	2.38
β_{ES}	.017	.019	.007	−.006	.023
ES at 5, 10, and 15 years	−.23:−.15:−.07	−.19:−.10:.00	−.15:−.11:−.07	.04:.01:−.02	−.25:−.13:−.02
Gender (k = 13) (d)[a]	−.09	−.12	−.04	.09	−.01
Girls	−.04	−.02	−.08*	−.04	.01
Boys	−.14**	−.12*	−.12**	.05	−.02
% ethnic minority ($Q_{\text{regression}(1)}$)	1.89	1.10	6.62*	3.51	12.19***
β_{ES}	−.0034	.0026	−.0082	−.0064	.0124
ES at 25%, 50%, and 75%	−.13:−.22:−.30	−.13:−.07:.00	−.04:−.24:−.43	.06:−.10:−.25	−.25:.06:.37
Country ($Q_{b(1)}$)[b]	0.35	0.06	0.74	0.96	2.68
United States (k = 17)	−.18***	−.11*	−.16***	−.03	−.09
Other countries (k = 6)	−.13*	−.12*	−.03	.12	−.31*
First author gender ($Q_{b(1)}$)	0.01	0.02	0.05	0.44	1.40
Female (k = 18)	−.17**	−.11**	−.11*	.00	−.13*
Male (k = 5)	−.16*	−.12*	−.13*	.07	−.27*
Gender in title ($Q_{b(1)}$)[c]	0.99	1.05	4.11*	3.32	7.10**
No (k = 10)	−.13*	−.15**	−.01	.12	−.31***
Yes (k = 13)	−.20***	−.08*	−.20***	−.07	−.03

Note. Significant differences among reporters from follow-up comparisons are denoted by different alphabetic subscripts. ES = effect size; CI = confidence interval.

[a] See Equation 3 and accompanying text for description of how gender moderation was evaluated. [b] Due to low representation of individual countries outside of the United States, this variable was dichotomized to studies conducted within versus outside of the United States. [c] Dichotomous variable of whether the word *gender* (or similar terms such as *sex* or *boys* and *girls*) was in title.

$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Discussion

Although developmental researchers have long understood the importance of studying childhood aggression, only recently has attention focused on the indirect forms this aggression can take. This meta-analysis synthesizes the available evidence to address three topics important to this focus: gender differences in direct and indirect aggression, the intercorrelation of these two forms, and the associations these two forms of aggression have with adjustment. Below, we discuss the findings of this meta-analysis, identify shortcomings of both this meta-analysis and the extant empirical literature, and suggest directions for future research.

Findings From the Current Meta-Analysis

Results regarding overall gender differences are consistent with prior reviews of the literature: Boys tend to enact more direct aggression than girls, but there is little gender difference in indirect aggression (our finding that girls enact more indirect aggression than boys was statistically significant but trivial in magnitude). Given the number of studies included here and in Archer's (2004) meta-analyses, the support for negligible gender differences in indirect aggression seems conclusive. This conclusion of trivial gender differences in indirect aggression challenges common portrayals of this form being more commonly enacted by girls than boys (e.g., Björkqvist

Peer Acceptance			Peer Rejection		
Direct Aggression		Indirect Aggression	Direct Aggression		Indirect Aggression
	.7			.7	
	.6		410	.6	2
	.5	7	630	.5	25678
	.4		66530000	.4	0578
	.3		200	.3	1346
4	.2		6	.2	05
71	.1	039	700	.1	66
70	.0	0	1	.0	045
776	-.0	189		-.0	5
8873	-.1	1235578		-.1	
321	-.2	1247		-.2	
964100	-.3	47		-.3	
43	-.4	25		-.4	
	-.5			-.5	
-.17	$\bar{r}$	-.12	.39	$\bar{r}$	.35
136.38	$Q_{(22)}$	229.49	250.50	$Q_{(21)}$	290.84

Figure 6. Stem-and-leaf plot of associations of direct and indirect aggression with peer relations.

et al., 1992; Crick, 1997). Instead, indirect aggression can be considered a form that is more equitably enacted by girls and boys than direct aggression (although the medium magnitude of gender differences in direct aggression should alert us to the fact that direct aggression is also a problem among girls). This gender similarity in indirect aggression generally held across methods of assessing aggression. Although reporter moderation was found, such that girls are viewed by teachers and parents as more indirectly aggressive than boys, and boys view themselves as more indirectly aggressive than do girls (no gender differences emerged for peer nominations or ratings or for observations), these differences were generally small.

The trivial gender difference in indirect aggression was also consistent across age, ethnicity, and country (i.e., no evidence of moderation). The failure to find age trends in gender differences is interesting in that it runs counter to expectations that socializing forces exacerbate gender differences with development (e.g., Baillargeon et al., 2007). We conclude that the general pattern is of *similarities* rather than differences among boys' and girls' use of indirect aggression. This conclusion raises questions regarding *why* the misperception that girls are more indirectly aggressive than boys is so pervasive. Evidence suggests that early-developing gender schemas guide biases in processing social information consistent with this misconception (Giles & Heyman, 2005). Recent popular accounts of indirect aggression among girls also likely contributes to a dualistic focus on gender differences (rather than similarities; Horn, 2004) among teachers, parents, and even researchers. More broadly, our conclusion of gender similarity in indirect aggression supports arguments that studying aggression among both genders, rather than focusing exclusively on boys, is an important endeavor (Pepler, Madsen, Webster, & Levene, 2005; Putallaz & Bierman, 2004; Underwood, 2003).

The second goal of this review was to examine the magnitude of intercorrelation between direct and indirect aggression. On average, this correlation was very high, suggesting that these two forms are overlapping manifestations of childhood aggression. At the same time, this correlation is not perfect, supporting factor-analytic studies showing the separability of these two constructs (Björkqvist et al., 1992; Crick & Grotpeter, 1995; Grotpeter & Crick, 1996; Hart et al., 1998; Macgowan, Nash, & Fraser, 2002; Vaillancourt et al., 2003). The average (corrected) correlation (.76) implies that about half (57%) of the variance in these

Table 9

Summary of Meta-Analytic Results of Associations of Direct and Indirect Aggression With Peer Rejection

	Direct		Indirect		
	r	sr	r	sr	Differential d
Heterogeneity (Q; $df = 21$)	250.50***	140.17***	290.84***	157.68***	58.97***
Random-effects mean ES	.39***	.20***	.35***	.09**	.10*
95% CI	0.32:0.46	0.14:0.26	0.27:0.43	0.03:0.15	0.02:0.18
Moderators					
Reporter ($Q_{b(3)}$)	41.49***	12.58**	33.60***	5.48	5.93
Observation ($k = 3$)	.17**$_{a}$	.16**$_{a}$	.10$_{a}$	.05	.14
Parent ($k = 1$)	.64***$_{c}$	.42***$_{c}$	.52***$_{d}$	.23***	.30*
Peer nomination ($k = 17$)	.41***$_{b}$	.22***$_{ab}$	.39***$_{c}$	.10*	.11*
Peer ratings ($k = 0$)	—	—	—	—	—
Self-report ($k = 0$)	—	—	—	—	—
Teacher ($k = 6$)	.36***$_{b}$	.34***$_{bc}$	.19*$_{b}$	−.03	.43**
Age ($Q_{\text{regression}(1)}$)	0.03	3.83	0.94	3.29	9.14**
β_{ES}	−.002	−.017	.013	.017	−.037
ES at 5, 10, and 15 years	.40:.39:.38	.27:.19:.10	.31:.36:.42	.02:.10:.19	.26:.07:−.11
Gender ($k = 10$) (d)[a]	.08	.06	.06	−.04	.00
Girls	.21**	.06	.23**	.11*	.00
Boys	.31***	.12*	.29***	.07	.00
% ethnic minority ($Q_{\text{regression}(1)}$)	1.17	1.29	1.21	0.08	0.86
β_{ES}	.0038	.0029	.0042	−.0008	.0033
ES at 25%, 50%, and 75%	.37:.45:.52	.17:.24:.31	.33:.42:.50	.11:.09:.07	.04:.13:.21
Country ($Q_{b(1)}$)[b]	4.78*	0.96	5.62*	0.43	0.01
United States ($k = 18$)	.42***	.22***	.38***	.10*	.10*
Other countries ($k = 4$)	.26***	.14*	.23***	.05	.09
First author gender ($Q_{b(1)}$)	0.01	2.63	0.94	4.55*	16.66***
Female ($k = 16$)	.39***	.16***	.38***	.14***	.01
Male ($k = 6$)	.40***	.32***	.26**	−.10	.49***
Gender in title ($Q_{b(1)}$)[c]	1.59	0.06	1.96	1.46	0.31
No ($k = 11$)	.35***	.21***	.29***	.05	.12*
Yes ($k = 11$)	.44***	.20***	.41***	.13**	.08

Note. Significant differences among reporters from follow-up comparisons are denoted by different alphabetic subscripts. ES = effect size; CI = confidence interval.

[a] See Equation 3 and accompanying text for description of how gender moderation was evaluated. [b] Due to low representation of individual countries outside of the United States, this variable was dichotomized to studies conducted within versus outside of the United States. [c] Dichotomous variable of whether the word *gender* (or similar terms such as *sex* or *boys* and *girls*) was in title.

$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

two forms overlaps, whereas the remaining variance is unique. Therefore, researchers should consider both the overlapping and the unique aspects of direct and indirect aggression in future work and examining the antecedents of both aggressiveness in general and direct or indirect aggression specifically is clearly warranted.

The overlap between direct and indirect aggression varied across several moderators (i.e., age, information source, and gender). Observations yield more distinct measures of direct versus indirect aggression than do other reports. Here, distinguishing an aggressive act as direct versus indirect is not always clear for untrained reporters (e.g., teachers, peers), and the cleanest measures are obtained via trained observers. Alternatively, it may be more difficult to observe indirect forms of aggression during the limited sampling periods often used, so the truly high correlation between forms is missed when using this methodology. Given that observations of aggression are difficult and time consuming, future research should evaluate whether the high overlap that exists among other sources of information is due to fuzziness in classifying acts and, if so, how researchers can obtain clearer assessments using other reporters. Although our review indicates which methods yield lower or higher correlations between direct and indirect aggression, it does not address which method yields

Table 10
Summary of Analyses Evaluating Publication Bias

	Published[a]	Publication quality[b]				
	$Q_{(1)}$	$Q_{(reg)}$	B_{ES}	$r_{N,\,ES}$[c]	Failsafe N[d]	Failsafe N_{min}[e]
Gender differences ($k = 107$)						
Direct	0.04	0.17	−.01	−.13	99,408	545
Indirect	1.14	0.90	−.01	.03	1,123	545
Intercorrelation ($k = 98$)	2.64	3.83	−.06	.02	565,533	500
Internalizing problems ($k = 27$)						
Direct *r*	3.01	1.80	−.05	.23		
Direct *sr*	1.64	0.35	−.01	.17		
Indirect *r*	1.64	1.44	−.05	.24	1,187	145
Indirect *sr*	0.13	0.37	−.02	.18	699	145
Differential *d*	0.74	0.07	−.01	−.03	925	145
Emotional dysregulation ($k = 18$)						
Direct *r*	3.00	2.41	−.09	−.16	6,459	100
Direct *sr*	1.61	0.44	−.03	.00	2,407	100
Indirect *r*	0.91	1.44	−.07	−.21	3,823	100
Indirect *sr*	0.06	0.64	−.04	−.22		
Differential *d*	2.09	0.16	−.03	.05	4,216	100
Delinquent behaviors ($k = 14$)						
Direct *r*	0.05	0.47	.04	−.24	5,477	80
Direct *sr*	1.20	2.25	.06	−.12	1,237	80
Indirect *r*	1.43	0.34	−.03	−.13	3,793	80
Indirect *sr*	2.10	1.96	−.04	−.05	216	80
Differential *d*	2.50	3.78	.14	−.08	1,328	80
Prosocial behavior ($k = 31$)						
Direct *r*	9.42**	7.60**	.08	.05	5,537	165
Direct *sr*	5.38*	5.35*	.08	.14	3,478	165
Indirect *r*	6.09*	4.48*	.06	−.06	2,058	165
Indirect *sr*	0.37	0.49	−.02	−.14	61	165
Differential *d*	1.92	5.21*	.10	.13	9,609	165
Peer acceptance ($k = 23$)						
Direct *r*	3.80	2.41	.04	−.01	888	125
Direct *sr*	1.27	1.28	.03	.06	349	125
Indirect *r*	0.65	0.56	.03	−.05	452	125
Indirect *sr*	0.05	0.06	.00	−.04		
Differential *d*	1.86	2.40	.06	.04	609	125
Peer rejection ($k = 22$)						
Direct *r*	20.15***	5.72*	−.08	−.16	5,511	120
Direct *sr*	1.13	0.63	−.02	−.51*	1,220	120
Indirect *r*	28.36***	7.39**	−.10	.11	4,711	120
Indirect *sr*	3.48	3.05	−.05	.37	426	120
Differential *d*	1.93	1.41	.04	−.46*	325	120

[a] Comparison of unpublished versus published studies. Values that represent $Q_{(1)}$ are distributed as $\chi^2_{(1)}$ under the null hypothesis of no difference between unpublished and published studies. [b] Regression of effect sizes onto continuous publication quality variable. Values that represent $Q_{regression}$ are distributed as $\chi^2_{(1)}$ under the null hypothesis of no linear trend of effect sizes across publication quality. B_{ES} represents linear change in effect sizes with increasing publication quality. [c] Correlation between study sample sizes and effect sizes; expected to be zero when there is no publication bias. [d] Rosenthal's (1979) failsafe *N*, which represents the number of excluded studies with $ES = 0$ that would need to exist before the effect size would be reduced to nonsignificance. [e] Rosenthal's (1979) recommended minimum failsafe number, $5k + 10$. This number is used to interpret Rosenthal's failsafe *N*. If Rosenthal's failsafe *N* is larger than this number, results can be considered robust to the threat of excluded studies yielding a nonsignificant effect.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

the "true" overlap between them. Perceptions of individuals not specifically trained to distinguish direct and indirect aggression (i.e., self-reports, peers, teachers) may be "blurred," resulting in inflated estimates of this overlap. Alternatively, the limited sampling of aggressive acts witnessed in observational studies may deflate the true overlap of children's tendencies to enact both forms of aggression. Multimethod research (using multiple sources of information in different contexts) may help answer this question. Further research on factors that influence perceptions of behaviors and individuals as directly and indirectly aggressive (e.g., Card, Hodges, Little, & Hawley, 2005; Neese, 1997; Ostrov, Crick, & Keating, 2005) will also aid in understanding the sources of this correlation.

Our results also indicate greater overlap of direct and indirect aggression among boys than girls. The combination of gender differences favoring boys' use of direct aggression and the trivial gender differences in indirect aggression (discussed above) suggests that both direct and indirect types are common in boys' behavioral repertoire. Because either form of aggression is common among boys, they may view both as possible ways to aggress against potential victims. Therefore, boys who enact one form of aggression may be inclined to also enact other forms of aggression, more so than for girls. The underlying reason for this gender moderation, however, is unclear. For example, biological or social forces may jointly promote both forms of aggression among boys, but one form or the other among girls. We might also ask whether these differences are veridical or a function of differential measurement processes across gender.

Arguably, the most important set of findings from this meta-analysis are the associations with maladjustment. Our findings indicate that direct and indirect aggression differentially relate to various aspect of maladjustment. Specifically, and consistent with our expectations, direct aggression is more strongly, and uniquely, associated with emotional dysregulation, conduct problems, low peer acceptance, and peer rejection (although indirect aggression also had unique associations with conduct problems and peer rejection). In contrast, indirect aggression is more strongly and uniquely associated with internalizing problems, perhaps reflecting that children with such problems are more willing to engage in these indirect, covert means of aggression than direct, confrontational forms. Interestingly, direct aggression exhibits a unique association with low prosocial behavior, whereas indirect aggression has a unique association with *high* prosocial behavior. Children using indirect aggression, which often requires the involvement of other peers (e.g., to aid in exclusion or rumor spreading), must also use prosocial skills to garner the support and assistance of others. Given the high correlation between direct and indirect aggression, most individual studies would not have sufficient power to identify these independent and differential associations. Here, the statistical power of meta-analytically combining results from multiple studies elucidates these unique links between the form of aggression and the maladjustment.

We found some evidence of moderation of these links, although caution should be exercised in drawing conclusions given the potential confounds among moderators among the limited number of studies considered (Lipsey, 2003). Method of assessing aggression was a common moderator with parent, teacher, and peer reports (nominations and ratings) of aggression having the strongest associations with maladjustment. On the one hand, observable aggression consistently enacted may be a strong reflection of maladjustment. On the other, these observers may be attuned to children's maladjustment, which makes them more aware of (or biased in their perceptions of) children's aggressive behavior. Further studies using multiple informants of both forms of aggression and adjustment would help clarify these possibilities.

Interestingly, we found no evidence of age or gender moderation. Although one must be cautious not to over interpret null findings, the absence of age differences suggests that the associations we found generally hold across childhood and adolescence. In other words, we have no reason to believe that either form of aggression is *only* linked to maladjustment during certain periods of development (countering our expectations regarding age normative or non-normative forms being differentially related to maladjustment). The absence of gender moderation of these effects is particularly informative in that it directly counters arguments that so-called gender nonnormative forms of aggression should be related to greater maladjustment (Crick, 1997). The lack of support for this hypothesis makes sense if one considers the trivial magnitude of gender differences in indirect aggression—this form is no more normative for one gender than the other given that boys and girls enact roughly equal levels of indirect aggression.

Limitations in Our Knowledge and Directions for Future Research

An inherent limitation of this study, as with all meta-analyses, is that the quality of conclusions must rely on the literature available. That is, a central concern in any meta-analysis is that the conclusions are only valid to the extent to which the included studies represent the population of research. Our reliance on random-effects (and mixed-effects) models allows for greater generalizability of findings (more so than if we had relied on fixed-effects models; Hedges & Vevea, 1998; Overton, 1998). Nevertheless, these models still assume that the included studies are a random sampling of the population of studies; to the extent that they are not, our results may be biased. Along these lines, two particular limits of this meta-analytic review merit consideration.

First, our review considered only studies investigating both direct and indirect forms of aggression. Although this limited sampling met our goal of comparing these two forms, it does raise questions of whether our findings can be generalized to either form individually. We identified 13 studies that investigated the effects considered in this review using indirect aggression only (i.e., not assessing or reporting on direct aggression). Supplemental analyses indicated that including these additional studies do not alter any of the conclusions reached here (i.e., tests of main effects and moderators yield identical patterns of significance and virtually identical effect sizes). More difficult to evaluate is whether inclusion of the thousands of studies investigating direct aggression only (until recently, the only form of aggression considered) would have affected our conclusions regarding direct aggression. We also performed supplemental analyses in which we included a small number of these studies and did not find any evidence that these studies changed our results. These supplemental analyses provide evidence that our sample of studies of indirect *and* direct aggression can be generalized to the population of studies measuring indirect *or* direct aggression. Nevertheless, it is possible that the large body of research investigating only direct aggression differs from those included in our review, so appropriate caution should be placed in generalizing these results.

A second consideration is whether our obtained sample of studies was less likely to include some findings than others (e.g., nonsignificant results). This threat, commonly called the file drawer problem (Rosenthal, 1979), is that researchers' file drawers are filled with studies of nonsignificant (or otherwise different) results. This threat is also considered within the general term of *publication bias* to reflect the possibility that nonsignificant or otherwise unacceptable results are less likely to be published (and therefore less likely to be included in meta-analyses) than studies finding significant or commonly accepted results. Although this threat can never be ruled out, we believe that it is unlikely to have biased our results for several reasons. First, our search procedures included several ways of locating unpublished studies (i.e., searching dissertation databases, soliciting unpublished studies from researchers in the field). Second, our analyses of publication biases yielded little support that unpublished studies produced different effect sizes than published studies or that there was censoring of low effect sizes among smaller studies. Third, failsafe numbers indicated that an enormous number of studies with null results would have to exist to invalidate our results. In short, we view it as unlikely file drawers are teeming with enough studies discrepant from those included so as to invalidate our results.

A more viable limitation of our sample of studies is the adequacy of testing moderators of effect sizes. In contrast to the ample body of research from which to analyze main effects, analysis of moderators of these effects (e.g., method of assessing aggression, age, gender) were often hampered by incomplete coverage across levels of these moderators. Moreover, other aspects of the studies may have accounted for these moderator effects. For example, the method of assessing aggression is confounded by the context in which the aggression might occur (e.g., observational measures are necessarily limited to the context of observation, teacher and peer reports capture primarily in-school aggression, parent reports might capture more aggression in the home or neighborhood, and self-reports might be nonspecific to context). Similarly, the moderators examined here might themselves be confounded (e.g., researchers tend to rely more on teacher or parent reports among younger children but self- and peer reports among older children and adolescents). The question of whether reporter, age, gender, or contextual features were the "true" moderators, or simply correlated with other, unmeasured study features, is difficult to answer within a meta-analysis (Lipsey, 2003). We therefore urge caution in interpreting these moderator results, and encourage future primary work to include wider age ranges, multiple reports of aggression in multiple contexts, systematically evaluate gender moderation, and compare results across cultures and nations in order to fully evaluate these moderating effects.

Our review of the extant literature has also identified some important limitations in this area of research. First, it is critical to consider the contexts in which most studies in our review measured aggression: Most used samples drawn from schools; therefore, many of the conclusions of this meta-analysis refer primarily to aggression within the school context. Furthermore, the primary literature suffers from both a lack of variability and inadequacy in reporting important contextual features, such as the size of groups in which children interact, the gender composition of these groups, the familiarity of individuals within and between these groups, and the settings (e.g., playground, classroom) and activities (e.g., structured vs. unstructured) in which aggression occurs. Future studies should closely consider, and compare, these contextual features.

A substantial limitation in our meta-analysis and the majority of research reviewed is the reliance on concurrent data. This problem is especially relevant to our understanding of the associations of direct and indirect aggression with maladjustment. As such, our analyses speak to the magnitude of association but say nothing about the direction of influence. Each aspect of maladjustment that we considered could be conceptualized as an antecedent or a consequence of direct and indirect aggression or simply as a correlation arising from common third-variable causes. Unfortunately, longitudinal investigations that evaluate these directions of influence are rare and those that do exist are so varied in terms of time span and other methodological features that meaningful meta-analytic combination is not yet possible. We strongly urge further experimental and longitudinal research using well-planned time spans (i.e., days or years, depending on presumed time span of effects), appropriate analytic strategies (at a minimum, controlling for initial levels of the presumed consequence), and wider sampling of ages and measurement strategies within these longitudinal investigations (Card & Little, 2007).

Finally, our meta-analytic review focused on the presence and magnitude of associations but did not evaluate the source of these associations. This criticism can also be leveled at most studies within the field. In addition to asking *whether* such associations exist, it is also important to ask *why* such associations exist. As indicated, longitudinal research begins to get at this question. However, a complete answer must recognize the mediating mechanisms among these effects, with emphasis on the shared versus unique mechanisms of direct versus indirect aggression. Similarly, it is important to understand when or *under what conditions* these associations exist (i.e., moderating processes). Although our analyses provided some evidence of these moderators, more work in the form of primary studies is needed to better understand the associations documented in this review.

Conclusions

Despite these limitations, this meta-analytic review does much to organize the existing literature so as to draw conclusions and guide future work. This work expands prior reviews of gender differences in aggression and is the first to examine (a) the intercorrelation between direct and indirect forms and (b) the associations these two forms have with maladjustment. In contrast to prior reviews showing that boys enact more direct aggression than girls (Hyde, 1984; Maccoby & Jacklin, 1974), our results (replicating those of Archer, 2004) generally demonstrate an absence of meaningful gender differences in indirect aggression. Although it can be argued that some methods of assessing indirect aggression yield some evidence of gender differences (also found by Archer, 2004), the magnitudes of these gender differences are uniformly small. We conclude that indirect aggression is *not* a "female form" of aggression.

Our finding that direct and indirect aggression shares about half of their variance has two implications. First, it highlights the need to consider commonalities between the two forms. Clearly, there are causes that explain children's aggression irrespective of form, and we should not lose sight of these commonalities in our quest to understand the differences. At the same time, there are differences, which are evident through the imperfect correlation between these two forms and our findings that direct and indirect aggression differentially relate to several aspects of maladjustment.

This meta-analytic review synthesizes the existing research on direct and indirect aggression during childhood and adolescence. This synthesis has resolved inconsistencies and misconceptions in the literature to reveal important information regarding gender similarities and differences, the commonality yet separability of these two forms, and their unique associations with maladjustment. However, the greater value of this work is in pointing to directions for future research. Clearly, research on these two forms of aggression has taken us well beyond the historic, narrow focus on boys' direct aggression. Just as clearly, more research on the indirect forms of aggression, among both boys and girls, is needed. This research should diverge from much of the prior work, however, to consider both the similarities and the differences in direct and indirect aggression and in recognizing that both forms are enacted by—and have similar correlates with maladjustment for—both genders.

References

References marked with an asterisk indicate studies included in the meta-analysis.

*Andreou, E. (2006). Social preference, perceived popularity and social intelligence: Relations to overt and relational aggression. *School Psychology International*, *27*, 339–351.

Archer, J. (2004). Sex differences in aggression in real-world settings: A meta-analytic review. *Review of General Psychology, 8*, 291 – 322.

Archer, J., & Coyne, S. M. (2005). An integrated review of indirect, relational, and social aggression. *Personality and Social Psychology Review, 9*, 212 – 230.

*Arnold, M. E. (1998). *Inflated sense of peer acceptance and hostile intent attributions in overtly and relationally aggressive children*. Unpublished dissertation, Texas A&M University.

Artz, S. (2005). The development of aggressive behaviors among girls: Measurement issues, social functions, and differential trajectories. In D. J. Pepler, K. C. Madsen, C. Webster, & K. S. Levene (Eds.), *The development and treatment of girlhood aggression* (pp. 105 – 136). Mahwah, NJ: Lawrence Erlbaum Associates.

Baillargeon, R. H., Zoccolillo, M., Keenan, K., Côté, S., Pérusse, D., Wu, H.-X., et al. (2007). Gender differences in physical aggression: A prospective population-based survey of children before and after 2 years of age. *Developmental Psychology, 43*, 13 – 26.

Barkley, R. A. (1997). Attention-deficit/hyperactivity disorder. In E. J. Mash & L. G. Terdal (Eds.), *Assessment of childhood disorders* (3rd ed., pp. 71 – 129). New York: Guilford.

*Barton, B. K., & Cohen, R. (2004). Classroom gender composition and children's peer relations. *Child Study Journal, 34*, 29 – 45.

Becker, B. J. (2005). Failsafe *N* or file-drawer number. In H. R. Rothsetin, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis Prevention, assessment and adjustments* (pp. 111 – 125). Hoboken, NJ: Wiley.

Begg, C. B. (1994). Publication bias. In H. Cooper & L. V. Hedges (Eds.), *Handbook of research synthesis* (pp. 399 – 409). New York: Russell Sage Foundation.

*Berdugo-Arstark, H. (2002). *The relationship of negative emotionality, hostile attribution bias, and social orientation to overt and relational aggression*. Unpublished dissertation, Fordham University.

Bjorklund, D. J., & Pellegrini, A. D. (2000). Child development and evolutionary psychology. *Child Development, 71*, 1687 – 1708.

Björkqvist, K. (1994). Sex differences in physical, verbal, and indirect aggression: A review of recent research. *Sex Roles, 30*, 177 – 188.

Björkqvist, K. (2001). Different names, same issue. *Social Development, 10*, 272 – 274.

*Björkqvist, K., Lagerspetz, K. M. J., & Kaukiainen, A. (1992). Do girls manipulate and boys fight? Developmental trends in regard to direct and indirect aggression. *Aggressive Behavior, 18*, 117 – 127.

*Blachman, D. R. (2003). *Predictors of peer rejection, acceptance, and victimization among girls with and without ADHD*. Unpublished dissertation, University of California, Berkeley.

*Bonica, C., Arnold, D. H., Fisher, P. H., Zeljo, A., & Yershova, K. (2003). Relational aggression, relational victimization, and language development in preschoolers. *Social Development, 12*, 551 – 562.

*Bosacki, S. L. (2003). Psychological pragmatics in preadolescents: Sociomoral understanding, self-worth, and school behavior. *Journal of Youth and Adolescence, 32*, 141 – 155.

*Brendgen, M., Dionne, G., Girard, A., Boivin, M., Vitaro, F., & Pérusse, D. (2005). Examining genetic and environmental effects on social aggression: A study of 6-year-old twins. *Child Development, 76*, 903 – 946.

Buss, A. H. (1961). *The psychology of aggression*. New York: Wiley.

*Butovskaya, M. L., Timentschik, V. M., & Burkova, V. N. (2007). Aggression, conflict resolution, popularity, and attitude to school in Russian adolescents. *Aggressive Behavior, 33*, 170 – 183.

*Cairns, R. B., Cairns, B. D., Neckerman, H. J., Ferguson, L. L., & Gariépy, J.-L. (1989). Growth and aggression: 1. Childhood to early adolescence. *Developmental Psychology, 25*, 320 – 330.

*Campbell, J. J. (1999). *Familial antecedents to children's relational and overt aggression*. Unpublished dissertation, University of North Carolina, Greensboro.

Card, N. A. (2003, April). Victims of peer aggression: A meta-analytic review. In N. A. Card & A. Nishina (Chairs), *Whipping boys and other victims of peer aggression: 25 years of research, now where do we go?* Poster symposium presented at the biennial meeting of the Society for Research in Child Development, Tampa, FL.

*Card, N. A., Hodges, E. V. E., Little, T. D., & Hawley, P. H. (2005). Gender effects in peer nominations for aggression and social status. *International Journal of Behavioral Development, 29*, 146 – 155.

Card, N. A., Isaacs, J., & Hodges, E. V. E. (2007). Correlates of school victimization: Recommendations for prevention and intervention. In J. E. Zins, M. J. Elias, & C. A. Maher (Eds.), *Bullying, victimization, and peer harassment: A handbook of prevention and intervention* (pp. 339 – 366). New York: Haworth Press.

Card, N. A., & Little, T. D. (2006). Proactive and reactive aggression in childhood and adolescence: A meta-analysis of differential relations with psychosocial adjustment. *International Journal of Behavioral Development, 30*, 466 – 480.

Card, N. A., & Little, T. D. (2007). Longitudinal modeling of developmental processes. *International Journal of Behavioral Development, 31*, 297 – 302.

*Carpenter, E. M. (2002). *A curriculum-based approach for social-cognitive skills training: An intervention targeting aggression in head start preschoolers*. Unpublished dissertation, University of Maine.

*Carpenter, E. M., & Nangle, D. W. (2006). Caught between stages: Relational aggression emerging as a developmental advance in at-risk preschoolers. *Journal of Research in Childhood Education, 21*, 177 – 188.

*Cillessen, A. H. N., Jiang, X. L., West, T. V., & Laszkowski, D. K. (2005). Predictors of dyadic friendship quality in adolescence. *International Journal of Behavioral Development, 29*, 165 – 172.

*Cillessen, A. H. N., & Mayeux, L. (2004). From censure to reinforcement: Developmental changes in the association between aggression and social status. *Child Development, 75*, 147–163.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Coie, J. D., & Dodge, K. A. (1998). Aggression and antisocial behavior. In W. Damon (Series Ed.) & N. Eisenberg (Vol. Ed.), *Handbook of child psychology* (Vol. 3, pp. 779–862). New York: Wiley.

*Costanzo, P. R., Killeya-Jones, K., Schmid, L. A., Mack, A. T., & Golonka, M. M. (2007). *Peer related processes related to substance use prevention in the middle to high school years.* [Data file]. An ongoing project of the Duke-NIDA Transdisciplinary Prevention Center supported by the National Institute on Drug Abuse Grant 1 P20 DA017589-01.

*Côté, S. M., Vaillancourt, T., Barker, E. D., Nagin, D., & Tremblay, R. E. (2007). The joint development of physical and indirect aggression: Predictors of continuity and change during childhood. *Development and Psychopathology, 19*, 37–55.

*Coyne, S. M., & Archer, J. (2005). The relationship between indirect and physical aggression on television and in real life. *Social Development, 14*, 324–338.

*Craig, W. M. (1998). The relationship among bullying, victimization, depression anxiety, and aggression in elementary school children. *Personality and Individual Differences, 24*, 123–130.

*Crain, M. M. (2002). *The relationship of intent attributions, goals and outcome expectancies to relationally aggressive behavior in pre-adolescent girls*. Unpublished dissertation, Alliant International University.

*Crick, N. R. (1995). Relational aggression: The role of intent attributions, feelings of distress, and provocation type. *Development and Psychopathology, 7*, 313–322.

*Crick, N. R. (1996). The role of overt aggression, relational aggression, and prosocial behavior in the prediction of children's future social adjustment. *Child Development, 67*, 2317–2327.

*Crick, N. R. (1997). Engagement in gender normative versus nonnormative forms of aggression: Links to social-psychological adjustment. *Developmental Psychology, 33*, 610–617.

Crick, N. R., Bigbee, M. A., & Howes, C. (1996). Gender differences in children's normative beliefs about aggression: How do I hurt thee? Let me count the ways. *Child Development, 67*, 1003–1014.

*Crick, N. R., Casas, J. F., & Mosher, M. (1997). Relational and overt aggression in preschool. *Developmental Psychology, 33*, 579–588.

*Crick, N. R., & Grotpeter, J. K. (1995). Relational aggression, gender, and social-psychological adjustment. *Child Development, 66*, 710–722.

*Crick, N. R., Grotpeter, J. K., & Bigbee, M. A. (2002). Relationally and physically aggressive children's intent attributions and feelings of distress for relational and instrumental peer provocations. *Child Development, 73*, 1134–1142.

*Crick, N. R., Ostrov, J. M., Burr, J. E., Cullerton-Sen, C., Jansen-Yeh, E., & Ralston, P. (2006). A longitudinal study of relational and physical aggression in preschool. *Applied Developmental Psychology, 27*, 254–268.

*Crick, N. R., Ostrov, J. M., & Werner, N. E. (2006). A longitudinal study of relational aggression, physical aggression, and children's social-psychological adjustment. *Journal of Abnormal Child Psychology, 34*, 131–142.

*Crick, N. R., & Werner, N. E. (1998). Response decision processes in relational and overt aggression. *Child Development, 69*, 1630–1639.

*Cristina, S. J. (2000). *Gossip and social exclusion in females: Do they have positive or negative consequences for social behavior?* Unpublished dissertation, University of Ottawa.

*Cunningham, C. E., Cunnigham, L. J., Martorelli, V., Tran, A., Young, J., & Zacharias, R. (1998). The effects of primary division, student-mediated conflict resolution programs on playground aggression. *Journal of Child Psychology and Psychiatry, 39*, 653–662.

*Dane, A. V. (2001). *A multimethod examination of the friendships of overtly aggressive and relationally aggressive children*. Unpublished dissertation, University of Toronto.

*David, C. F. (2001). *A longitudinal investigation of the causal relationship between positive perceptual bias and interpersonal aggression among children*. Unpublished dissertation, Florida State University.

*David, C. F., & Kistner, J. A. (2000). Do positive self-perceptions have a "dark side"? Examination of the link between perceptual bias and aggression. *Journal of Abnormal Child Psychology, 28*, 327–337.

*Delveaux, K. (2003). *Peer groups and social adjustment in adolescence: A relational approach to the study of relational aggression*. Unpublished dissertation, Carleton University.

*DeRosier, M. E., & Thomas, J. M. (2003). Strengthening sociometric prediction: Scientific advances in the assessment of children's peer relations. *Child Development, 75*, 1379–1392.

*Dettling, A. C., Gunnar, M. R., & Donzella, B. (1999). Cortisol levels of young children in full-day childcare centers: Relations with age and temperament. *Psychoneuroendocrinology, 24*, 519–536.

*Estrem, T. L. (2003). *Relational and physical aggression among preschoolers: The effect of language skills and gender.* Unpublished dissertation, University of Minnesota.

*Estrem, T. L. (2005). Relational and physical aggression among preschoolers: The effect of language skills and gender. *Early Education & Development, 16*, 209–231.

*Evans, M. A. (2005). *Links between personality traits and school aggression and internalizing behaviors in African American early adolescents*. Unpublished dissertation, Syracuse University.

*Feshbach, N. D. (1969). Sex differences in children's modes of aggressive responses toward outsiders. *Merrill-Palmer Quarterly, 15*, 249–258.

Feshbach, S. (1970). Aggression. In P. H. Mussen (Ed.), *Carmichael's manual of child psychology* (Vol. 2, pp. 159–259). New York: Wiley.

*Finch, C. L. (2001). *The relationship among relationally aggressive behavior, emotion and social cognitions in preadolescent* *females*. Unpublished dissertation, Alliant International University.

*Fite, P. J., Stauffacher, K., Ostrov, J. M., & Colder, C. R. (2008). Replication and extension of Little et al.'s (2003) forms and functions of aggression measure. *International Journal of Behavioral Development, 32*, 238 – 242.

*Foo, G. B. (2002). *Marital conflict and childhood relational aggression: A gender-balanced approach to understanding adjustment problems in middle childhood*. Unpublished dissertation, University of Minnesota.

*Foster, H. A. (2001). *Neighborhood and family contexts of gendered aggression in childhood*. Unpublished dissertation, University of Toronto.

Galen, B. R., & Underwood, M. K. (1997). A developmental investigation of social aggression among children. *Developmental Psychology, 33*, 589 – 600.

*Geiger, T. C. (2003). *The influence of peer status, friendship, and children's social behavior on trajectories of children's peer victimization*. Unpublished dissertation, University of Minnesota.

Giles, J. W., & Heyman, G. D. (2005). Young children's beliefs about the relationship between gender and aggressive behavior. *Child Development*, *76*, 107 – 121.

*Gleason, K. A., Jensen-Campbell, L. A., & Richardson, D. S. (2004). Agreeableness as a predictor of aggression in adolescence. *Aggressive Behavior, 30*, 43 – 61.

*Goldstein, S. E. (2003). *Social reasoning about relational aggression during adolescence*. Unpublished dissertation, Bowling Green State University.

*Grotpeter, J. K. (1997). *Relational aggression, overt aggression, and family relationships*. Unpublished dissertation, University of Illinois.

Grotpeter, J. K., & Crick, N. R. (1996). Relational aggression, overt aggression, and friendship. *Child Development*, *67*, 2328 – 2338.

*Hart, C. H., Nelson, D. A., Robinson, C. C., Olsen, S. F., McNeilly-Choque, M. K. (1998). Overt and relational aggression in Russian nursery-school-age children: Parenting style and marital linkages. *Developmental Psychology, 34*, 687 – 697.

*Hawley, P. H. (2003). Strategies of control, aggression, and morality in preschoolers: An evolutionary perspective. *Journal of Experimental Child Psychology, 85*, 213 – 235.

*Hawley, P. H., Little, T. D., & Card, N. A. (2007). The allure of a mean friend: Relationship quality and processes of aggressive adolescents with prosocial skills. *International Journal of Behavioral Development, 31*, 170 – 180.

*Hawley, P. H., Little, T. D., & Card, N. A. (2008). The myth of the alpha male: A new look at dominance-related beliefs and behaviors among adolescent males and females. *International Journal of Behavioral Development, 32*, 76 – 88.

Hawley, P. H., Little, T. D., & Rodkin, P. C. (Eds.). (2007). *Aggression and adaptation*. Mahwah, NJ: Lawrence Erlbaum Associates.

Hays, W. L. (1994). *Statistics* (5th ed.). Belmont, CA: Wadsworth.

*Hayward, S. M., & Fletcher, J. (2003). Relational aggression in an Australian sample: Gender and age differences. *Australian Journal of Psychology, 55*, 129 – 134.

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods, 3*, 486 – 504.

*Hektner, J. M., & Swenson, C. (2005). Unpublished data.

*Henington, C. D. (1996). *Social correlates of relational and overt aggression in boys and girls*. Unpublished dissertation, Texas A & M University.

Henington, C., Hughes, J. N., Cavell, T. A., & Thompson, B. (1998). The role of relational aggression in identifying aggressive boys and girls. *Journal of School Psychology, 36*, 457 – 477.

Horn, S. S. (2004). Mean girls or cultural stereotypes? *Human Development*, *47*, 314 – 320.

*Hughes, J. H., Meehan, B. T., & Cavell, T. A. (2004). Development and validation of a gender-balanced measure of aggression-relevant social cognition. *Journal of Clinical Child and Adolescent Psychology, 33*, 292 – 302.

*Hunt, M. H. (2002). *Exploratory and confirmatory factor analysis of the student survey of bullying behavior*. Unpublished dissertation, Georgia State University.

Hunter, J. E., & Schmidt, F. L. (1990). Dichotomization of continuous variables: The implications for meta-analysis. *Journal of Applied Psychology, 75*, 334 – 349.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Newbury Park, CA: Sage.

Hyde, J. S. (1984). How large are gender differences in aggression? A developmental meta-analysis. *Developmental Psychology, 20*, 722 – 736.

Johnson, D. R. (2003). *The relationship between relational aggression in preschool children and friendship stability, mutuality, and popularity*. Unpublished dissertation, Alliant International University.

*Juliano, M., Werner, R. S., & Cassidy, K. W. (2006). Early correlates of preschool aggressive behavior according to type of aggression and measurement. *Journal of Applied Developmental Psychology, 27*, 395 – 410.

*Kaukiainen, A., Björkqvist, K., Lagerspetz, K. M. J., Österman, K., Salmivalli, C., Rothberg, S., et al. (1999). The relationships between social intelligence, empathy, and three types of aggression. *Aggressive Behavior, 25*, 81 – 89.

*Killeya-Jones, L. A., & Costanzo, P. R. (2008). *The role of aggression in peer evaluation in a racially-balanced sample of early adolescents*. Manuscript submitted for publication.

*Krause, M. M. (2005). *Teacher, peer, and self-assessment of aggression and perceptions of school attachment among elementary and secondary, regular education and learning disabled students*. Unpublished dissertation, Wayne State University.

Ladd, G. W. (2005). *Children's peer relations and social competence: A century of progress*. New Haven, CT: Yale University Press.

*Lafferty, J. (2003). *The relationship between gender, empathy, and aggressive behaviors among early adolescents*. Unpublished dissertation, Alliant International University.

*Lagerspetz, K. M. J., & Björkqvist, K. (1994). Indirect aggression in boys and girls. In L. R. Huesmann (Ed.), *Aggressive behavior: Current perspectives* (pp. 131–150). New York: Plenum Press.

*Lagerspetz, K. M. J., Björkqvist, K., & Peltonen, T. (1988). Is indirect aggression typical of females? Gender differences in aggressiveness in 11- to 12-year-old children. *Aggressive Behavior, 14*, 403–414.

*Lancelotta, G. X., & Vaughn, S. (1989). Relation between types of aggression and sociometric status: Peer and teacher perceptions. *Journal of Educational Psychology, 81*, 86–90.

*Landau, S. F., Björkqvist, K., Lagerspetz, K. M. J., Österman, K., & Gideon, L. (2002). The effect of religiosity and ethnic origin on direct and indirect aggression among males and females: Some Israeli findings. *Aggressive Behavior, 28*, 281–298.

*Lease, A. M., Kennedy, C. A., & Axelrod, J. L. (2002). Children's social constructions of popularity. *Social Development, 11*, 87–109.

*Leff, S. S. (1995). *Females' cognitions of their relationally aggressive peers as a function of sociometric and relationally aggressive status.* Unpublished dissertation, University of North Carolina, Chapel Hill.

Lipsey, M. W. (2003). Those confounded moderators in meta-analysis: Good, bad, and ugly. *The Annals, 587*, 69–81.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage.

*Little, T. D., Jones, S. M., Henrich, C. C., & Hawley, P. H. (2003). Disentangling the "whys" from the "whats" of aggressive behavior. *International Journal of Behavioral Development, 27*, 122–133.

*Lockwood, R. L. (2002). *Examination of siblings' aggression styles: Do sisters show more relational aggression than brothers?* Unpublished dissertation, University of Memphis.

*Long, M. M. (2003). *Investigation of relational and overt aggression among boys and girls.* Unpublished dissertation, University of North Texas.

*Loukas, A., Paulos, S. K., & Robinson, S. (2005). Early adolescent social and overt aggression: Examining the roles of social anxiety and maternal psychological control. *Journal of Youth and Adolescence, 34*, 335–345.

Maccoby, E. E. (1990). Gender and relationships: A developmental account. *American Psychologist, 45*, 513–520.

Maccoby, E. E. (1998). *The two sexes: Growing up apart, coming together.* Cambridge, MA: Harvard University Press.

Maccoby, E. E., & Jacklin, C. N. (1974). *The psychology of sex differences.* Stanford, CA: Stanford University Press.

*Macgowan, M. J., Nash, J. K., & Fraser, M. W. (2002). The Carolina Child Checklist of risk and protective factors for aggression. *Research on Social Work Practice, 12*, 253–276.

*McCay, L. A. (2004). *Outcome expectancy and value biases among overtly and relationally aggressive children.* Unpublished dissertation, State University of New York, Binghamtom.

*McEvoy, M. A., Estrem, T. L., Rodriguez, M. C., & Olsen, M. L. (2003). Assessing relational and physical aggression among preschool children: Intermethod agreement. *Topics in Early Childhood Special Education, 23*, 53–63.

McMahon, R. J., & Estes, A. M. (1997). Conduct problems. In E. J. Mash & L. G. Terdal (Eds.), *Assessment of childhood disorders* (3rd ed., pp. 130–193). New York: Guilford.

*McNeill, S. K. (2001). *The relation between parental discipline practices, gender, and children's overt and relational aggression.* Unpublished dissertation, University of Minnesota.

*McNeilly-Choque, M. K., Hart, C. H., Robinson, C. C., Nelson, L. J., & Olsen, S. F. (1996). Overt and relational aggression on the playground: Correspondence between different informants. *Journal of Research in Childhood Education, 11*, 47–67.

*Meadow, A. S. (2001). *Teacher, parent and student assessment of social, internalizing, and externalizing problem behaviors among three types of aggressive students.* Unpublished dissertation, Wayne State University.

*Miller, D. R. (2001). *Friendship and romantic relationship features among relationally aggressive adolescent girls.* Unpublished dissertation, California School of Professional Psychology.

*Miller, J. L., Vaillancourt, T., & Boyle, M. H. (in press). Examining the heterotypic continuity of aggression using teacher reports: Results from a national Canadian study. *Social Development.*

*Mizokawa, S. T. (1999). *Interparental conflict and child and adolescent aggression: An examination of overt and relational aggression.* Unpublished dissertation, University of Nebraska.

*Murray-Close, D. (2006). *Gender differences in the association between autonomic arousal and aggressive conduct.* Unpublished dissertation, University of Minnesota.

*Murray-Close, D., & Crick, N. R. (2006). Mutual antipathy involvement: Gender and associations with aggression and victimization. *School Psychology Review, 35*, 472–492.

*Murray-Close, D., Crick, N. R., & Galotti, K. M. (2006). Children's moral reasoning regarding physical and relational aggression. *Social Development, 15*, 345–372.

*Musher-Eizenman, D. R., Boxer, P., Danner, S., Dubow, E. F., Goldstein, S. E., Heretick, D. M. L. (2004). Social-cognitive mediators of the relation of environmental and emotion regulation factors to children's aggression. *Aggressive Behavior, 30*, 389–408.

Neese, D. E. K. (1997). *Teachers' perceptions of gender differences in child aggression.* Unpublished dissertation, Purdue University.

*Nelson, D. A., & Crick, N. R. (2002). Parental psychological control: Implications for childhood physical and relational aggression. In B. K. Barber (Ed.), *Intrusive parenting: How psychological control affects children and adolescents* (pp. 161–189). Washington, DC: American Psychological Association.

*Nelson, D. A., Hart, C. A., Yang, C., Olsen, J. A., & Jin, S. (2006). Aversive parenting in China: Associations with child physical and relational aggression. *Child Development, 77*, 554–572.

*Nelson, D. A., Robinson, C. C., & Hart, C. H. (2005). Relational and physical aggression of preschool-age children: Peer status linkages across informants. *Early Education and Development*, *16*, 115–139.

*O'Donnell, A. S. (2002). *Implementation and evaluation of a treatment program for relational aggression and victimization with preadolescent girls*. Unpublished dissertation, Alliant International University.

*Onyskiw, J. E. (1999). *Processes underlying children's responses to witnessing physical aggression in their families*. Unpublished dissertation, University of Alberta.

*Osantowski, J. D. (2001). *The social cognitions associated with relational aggression and depression in preadolescent girls*. Unpublished dissertation, California School of Professional Psychology.

*O'Shea, H. K. (2004). *Assessment of aggression in children: An exploration of self and peer report methods*. Unpublished dissertation, Alliant International University.

*Österman, K., Björkqvist, K., Lagerspetz, K. M. J., Kaukiainen, A., Huesmann, L. R., & Frączek, A. (1994). Peer and self-estimated aggression and victimization in 8-year-old children from five ethnic groups. *Aggressive Behavior*, *20*, 411–428.

*Ostrov, J. M. (2006). Deception and subtypes of aggression during early childhood. *Journal of Experimental Child Psychology*, *93*, 322–336.

*Ostrov, J. M. (2008). *Forms of aggression and peer victimization during early childhood: A short-term longitudinal study*. Manuscript submitted for publication.

*Ostrov, J. M., & Crick, N. R. (2007). Forms and functions of aggression during early childhood. A short-term longitudinal study. *School Psychology Review*, *36*, 22–43.

Ostrov, J. M., Crick, N. R., & Keating, C. F. (2005). Gender-biased perceptions of preschoolers' behavior: How much is aggression and prosocial behavior in the eye of the beholder? *Sex Roles*, *52*, 393–398.

*Ostrov, J. M., Crick, N. R., & Stauffacher, K. (2006). Relational aggression in sibling and peer relationships during early childhood. *Applied Developmental Psychology*, *27*, 241–253.

*Ostrov, J. M., & Keating, C. F. (2004). Gender differences in preschool aggression during free play and structured interactions: An observational study. *Social Development*, *13*, 255–277.

*Ostrov, J. M., Woods, K. E., Jansen, E. A., Casas, J. F., & Crick, N. R. (2004). An observational study of delivered and received aggression, gender, and social-psychological adjustment in preschool: "This white crayon doesn't work..." *Early Childhood Research Quarterly*, *19*, 355–371.

Overton, R. C. (1998). A comparison of fixed-effects and mixed (random-effects) models for meta-analysis tests of moderator variable effects. *Psychological Methods*, *3*, 354–379.

*Pakaslahti, L., & Keltikangas-Järvinen, L. (1998). Types of aggressive behavior among aggressive-preferred, aggressive non-preferred, non-aggressive preferred and non-aggressive non-preferred 14-year-old adolescents. *Personality and Individual Differences*, *24*, 821–828.

Pakaslahti, L., & Keltikangas-Järvinen, L. (2000). Comparison of peer, teacher, and self-assessments on adolescent direct and indirect aggression. *Emotional Psychology*, *20*, 177–190.

*Park, J.-H., Essex, M. J., Zahn-Waxler, C., Armstrong, J. M., Klein, M. H., & Goldsmith, H. H. (2005). Relational and overt aggression in middle childhood: Early child and family risk factors. *Early Education and Development*, *16*, 233–287.

Parke, R. D., & Slaby, R. G. (1983). The development of aggression. In P. H. Mussen (Ed.), *Handbook of child psychology* (Vol. 4, pp. 547–641). New York: Wiley.

*Peets, K., & Kikas, E. (2006). Aggressive strategies and victimization during adolescence: Grade and gender differences, and cross-informant agreement. *Aggressive Behavior*, *32*, 68–79.

*Pepler, D. J., Craig, W. M., & Roberts, W. L. (1998). Observations of aggressive and nonaggressive children on the school playground. *Merrill-Palmer Quarterly*, *44*, 55–76.

*Pepler, D. J., Madsen, K. C., Webster, C., & Levene, K. S. (Eds.). (2005). *The development and treatment of girlhood aggression*. Mahwah, NJ: Erlbaum.

*Phillipsen, L. C., Deptula, D. P., & Cohen, R. (1999). Relating characteristics of children and their friends to relational and overt aggression. *Child Study Journal*, *29*, 269–289.

*Prinstein, M. J., Boergers, J., & Vernberg, E. M. (2001). Overt and relational aggression in adolescents: Social-psychological adjustment of aggressors and victims. *Journal of Clinical Child Psychology*, *30*, 479–491.

Putallaz, M., & Bierman, K. L. (Eds.). (2004). *Aggression, antisocial behavior, and violence among girls: A developmental perspective*. New York: Guilford.

*Rana, M., & Malhotra, D. (2005). Family environment as a predictor of aggressive behavior. *Studia Psychologica*, *47*, 61–74.

*Rielly, N. E. (2003). *Girls and boys identified with a subclinical cutoff for attention problems: Behavioural, emotional and social characteristics*. Unpublished dissertation, Queen's University.

*Rockhill, C. M. (2000). *Depressed children's friendships*. Unpublished dissertation, University of Illinois.

Rodriguez, M. C., & Maeda, Y. (2006). Meta-analysis of coefficient alpha. *Psychological Methods*, *11*, 306–322.

*Rose, A. J., Swenson, L. P., & Waller, E. M. (2004). Overt and relational aggression and perceived popularity: Developmental differences in concurrent and prospective relations. *Developmental Psychology*, *40*, 378–387.

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin*, *86*, 638–641.

Rosenthal, R. (1991). *Meta-analytic procedures for social research* (Rev. ed.). Newbury Park, CA: Sage.

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (Eds.). (2005). *Publication bias in meta-analysis Prevention, assessment and adjustments*. Hoboken, NJ: Wiley.

*Russell, A., Hart, C. H., Robinson, C. C., & Olsen, S. F. (2003). Children's sociable and aggressive behaviour with peers: A comparison of the US and Australia, and contributions of temperament and parenting styles. *International Journal of Behavioral Development*, *27*, 74–86.

*Rys, G. S., & Bear, G. G. (1997). Relational aggression and peer relations: Gender and developmental issues. *Merrill-Palmer Quarterly, 43*, 87–106.

*Salmivalli, C., & Kaukiainen, A. (2004). "Female aggression" revisited: Variable- and person-centered approaches to studying gender differences in different types of aggression. *Aggressive Behavior, 30*, 158–163.

*Salmivalli, C., Kaukiainen, A., & Lagerspetz, K. (2000). Aggression and sociometric status among peers: Do gender and type of aggression matter? *Scandinavian Journal of Psychology, 41*, 17–24.

*Schmidt, M. G. (2004). *The roles of gender, aggression, and peer influence in the development of alcohol use in adolescence*. Unpublished dissertation, University of Virginia.

*Sebanc, A. M. (1999). *Friendship experiences among preschool children: Links with prosocial behavior and aggression*. Unpublished dissertation, University of Minnesota.

*Sebanc, A. M. (2003). The friendship features of preschool children: Links with prosocial behavior and aggression. *Social Development, 12*, 249–268.

*Selah-Shayovits, R. (2004). School for aggression: Types of adolescent aggression in school students and school dropouts. *International Journal of Adolescence and Youth, 11*, 303–316.

*Simon, L. (2001). *Psychosocial correlates of relational and overt aggression in preschool children*. Unpublished dissertation, Fordham University.

*Smith, C. E. (2004). *Naturalistic observations of girls' and boys' playground aggression*. Unpublished dissertation, York University.

*Smith, R. G. (2005). *Bullying: Prevalence, subtypes, developmental trends, and the effect of gender*. Unpublished dissertation, University of Mississippi.

*Smith, R. L., Rose, A. J., & Schwartz-Mette, R. A. (2008). *Relational and overt aggression in childhood and adolescence: Clarifying mean-level gender differences and associations with peer acceptance*. Manuscript submitted for publication.

*Solis, N. A. (1998). *Disruptive behavior disorders: Development of a DSM-IV-based self-report measure for adolescents*. Unpublished dissertation, University of Toledo.

*Suarez, M. (2001). *Relational aggression: An analysis of early childhood indicators*. Unpublished dissertation, Wichita State University.

*Sullivan, T. N., Farrell, A. D., & Kliewer, W. (2006). Peer victimization in early adolescence: Association between physical and relational victimization and drug use, aggression, and delinquent behaviors among urban middle school students. *Development and Psychopathology, 18*, 119–137.

*Tapper, K., & Boulton, M. J. (2004). Sex difference in levels of physical, verbal, and indirect aggression amongst primary school children and their associations with beliefs about aggression. *Aggressive Behavior, 30*, 125–145.

*Tiet, Q. Q., Wasserman, G. A., Loeber, R., McReynolds, L. S., & Miller, L. S. (2001). Developmental and sex differences in types of conduct problems. *Journal of Child and Family Studies, 10*, 181–197.

*Toldos, M. P. (2005). Sex and age differences in self-estimated physical, verbal, and indirect aggression in Spanish adolescents. *Aggressive Behavior, 31*, 13–23.

*Tomada, G., & Schneider, B. H. (1997). Relational aggression, gender, and peer acceptance: Invariance across culture, stability over time, and concordance among informants. *Developmental Psychology, 33*, 601–609.

Tremblay, R. E. (2000). The development of aggressive behavior during childhood: What have we learned in the past century? *International Journal of Behavioral Development, 24*, 129–141.

Trivers, R. (1972). Parental investment and sexual selection. In B. B. Campbell (Ed.), *Sexual selection and the descent of man* (pp. 136–179). Chicago: Alpine.

*Underwood, M. K. (2003). *Social aggression among girls*. New York: Guilford Press.

*Underwood, M. K., Galen, B. R., & Paquette, J. A. (2001). Top ten challenges for understanding gender and aggression in children: Why can't we all just get along? *Social Development, 10*, 248–266.

*Underwood, M. K., Scott, B. L., Galperin, M. B., Bjornstad, G. J., & Sexton, A. M. (2004). An observational study of social exclusion under varied conditions: Gender and developmental differences. *Child Development, 75*, 1538–1555.

*Vaillancourt, T. (2001). *Competing for hegemony during adolescence: A link between aggression and social status*. Unpublished dissertation, University of British Columbia.

*Vaillancourt, T., Brendgen, M., Boivin, M., & Tremblay, R. E. (2003). A longitudinal confirmatory factor analysis of indirect and physical aggression: Evidence of two factors over time? *Child Development, 74*, 1628–1638.

*Vaillancourt, T., & Hymel, S. (2006). Aggression and social status: The moderating roles of sex and peer-valued characteristics. *Aggressive Behavior, 32*, 396–408.

*Verlaan, P. (1995). *Gender-related aggressive strategies, psychosocial adjustment, and parental influences in middle childhood*. Unpublished dissertation, Concordia University.

*Weiner, M. D. (2002). *Aggression and victimization as predictors of drug use in early adolescence*. Unpublished dissertation, University of Southern California.

*Werner, N. E. (2000). *Friends' influence on changes in externalizing behavior during middle childhood: A longitudinal study of relational and physical aggression*. Unpublished dissertation, University of Minnesota.

*Werner, N. E., & Crick, N. R. (2004). Maladaptive peer relationships and the development of relational and physical aggression during middle childhood. *Social Development, 13*, 495–514.

*Werner, N. E., & Nixon, C. L. (2005). Normative beliefs and relational aggression: An investigation of the cognitive bases of adolescent aggressive behavior. *Journal of Youth and Adolescence, 34*, 229–243.

*Willoughby, M., Kupersmidt, J., & Bryant, D. (2001). Overt and covert dimensions of antisocial behavior in early childhood. *Journal of Abnormal Child Psychology, 29*, 177–187.

*Willoughby, T., Chalmers, H., & Busseri, M. A. (2004). Where is the problem? Examining co-occurrence among multiple problem behaviors in adolescence. *Journal of Consulting and Clinical Psychology, 72,* 1022–1037.

Wilson, M., & Daly, M. (1985). Competitiveness, risk taking, and violence: The young male syndrome. *Ethology and Sociobiology, 6,* 59–73.

*Wolke, D., Woods, S., Bloomfield, L., Karstadt, L. (2000). The association between direct and relational bullying and behavior problems among primary school children. *Journal of Child Psychology and Psychiatry, 41,* 989–1002.

*Woods, S., Wolke, D., Nowicki, S., & Hall, L. (2008). *Emotion recognition abilities and empathy of victims of bullying or peer rejection.* Manuscript submitted for publication.

*Xie, H., Cairns, B. D., & Cairns, R. B. (2002). The development of social aggression and physical aggression: A narrative analysis of interpersonal conflicts. *Aggressive Behavior, 28,* 341–355.

Xie, H., Cairns, B. D., & Cairns, R. B. (2005). The development of aggressive behaviors among girls: Measurement issues, social functions, and differential trajectories. In D. J. Pepler, K. C. Madsen, C. Webster, & K. S. Levene (Eds.), *The development and treatment of girlhood aggression* (pp. 105–136). Mahwah, NJ: Lawrence Erlbaum Associates.

*Xie, H., Farmer, T. W., & Cairns, B. D. (2003). Different forms of aggression among inner-city African-American children: Gender, configurations, and school social networks. *Journal of School Psychology, 41,* 355–375.

*Xie, H., Swift, D. J., Cairns, B. D., & Cairns, R. B. (2002). Aggressive behaviors in social interaction and developmental adaptation: A narrative analysis of interpersonal conflicts during early adolescence. *Social Development, 11,* 205–224.

*Zahn-Waxler, C., Park, J.-H., Essex, M., Slattery, M., & Cole, P. M. (2005). Relational and overt aggression in disruptive adolescents: Prediction from early social representations and links with concurrent problems. *Early Education & Development, 16,* 259–282.

*Zalecki, C. A., & Hinshaw, S. P. (2004). Overt and relational aggression in girls with attention deficit hyperactivity disorder. *Journal of Clinical Child and Adolescent Psychology, 33,* 125–137.

*Zimmer-Gembeck, M. J., Geiger, T. C., & Crick, N. R. (2005). Relational and physical aggression, prosocial behavior, and peer relations: Gender moderation and bidirectional associations. *Journal of Early Adolescence, 25,* 421–452.

*Zimmer-Gembeck, M. J., Hunter, T. A., & Pronk, R. (2007). A model of behaviors, peer relations and depression: Perceived social acceptance as a mediator and the divergence of perceptions. *Journal of Social and Clinical Psychology, 26,* 273–302.

Appendix

Table A1
Studies Reporting Gender Differences in Direct and Indirect Aggression

Study	*N*	Age (years)	Reporter	Publication status[a]	Author gender	Country	% Ethnic minority	Direct (*r*)	Indirect (*r*)
Barton and Cohen (2004)	39	10.0	Peer nominations	4	M	United States	44	−.01	−.13
Berdugo-Arstark (2002)	128	8.8	Teacher	2	F	United States	0	.12	−.17
Björkqvist et al. (1992)	85	8.5	Peer rating	4	M	Finland	—	.19	−.13
Bosacki (2003)	239	10.8	Teacher	4	F	Canada	1	.23	.00
Brendgen et al. (2005)	468	6.1	Peer nominations	5	F	Canada	8	.49	.15
Brendgen et al. (2005)	468	6.1	Teacher	5	F	Canada	8	.22	−.14
Butovskaya, Timentschik, and Burkova (2007)	212	13.7	Peer rating	4	F	Russia	—	.29	−.20
Butovskaya et al. (2007)	212	13.7	Self	4	F	Russia	—	.12	−.21
Cairns et al. (1989)	215	12.2	Other (victim)	5	M	United States	12	.27	−.21
Campbell (1999)	139	10.0	Teacher	2	F	United States	42	.18	−.03
Card et al. (2005)	351	10.5	Peer nominations	4	M	United States	32	.36	−.26
Carpenter and Nangle (2006)	82	4.0	Teacher	4	F	United States	6	.17	.04
Côté, Vaillancourt, Barker, Nagin, and Tremblay (2007)	1,183	3.4	Parent	5	F	Canada	17	.12	−.09
Coyne and Archer (2005)	347	12.5	Peer nominations	4	F	England	—	.25	−.25
Craig (1998)	546	11.2	Self	4	F	Canada	29	.13	−.03
Crick (1995)	252	9.4	Peer nominations	5	F	United States	27	.27	.11
Crick (1997)	1,166	10.5	Peer nominations	5	F	United States	16	.37	−.08
Crick and Grotpeter (1995)	491	9.4	Peer nominations	5	F	United States	40	.36	−.14

Table A1
Continued

Study	*N*	Age (years)	Reporter	Publication status[a]	Author gender	Country	% Ethnic minority	Direct (*r*)	Indirect (*r*)
Crick et al. (1996)	162	10.0	Peer nominations	5	F	United States	49	.17	−.32
Crick, Casas, and Mosher (1997)	65	4.5	Peer nominations	5	F	United States	27	.00	.00
Crick et al. (1997)	65	4.5	Teacher	5	F	United States	27	.39	−.36
Crick, Grotpeter, and Bigbee (2002)—Study 1	121	9.0	Peer nominations	5	F	United States	38	.62	−.48
Crick et al. (2002)—Study 2	485	10.5	Peer nominations	5	F	United States	28	.39	−.08
Crick, Ostrov, and Werner (2006)	234	8.0	Peer nominations	5	F	United States	40	.30	−.13
Crick, Ostrov, Burr et al. (2006)	91	4.0	Observation	4	F	United States	43	.48	−.28
David (2001); David and Kistner (2000)	749	9.8	Peer nominations	5	F	United States	31	.41	.12
Delveaux (2003)	425	12.5	Peer nominations	2	F	Canada	0	.39	−.28
Estrem (2003)	100	4.2	Teacher	2	F	United States	11	.37	.08
Evans (2005)	133	12.4	Self	2	F	United States	100	.00	.08
Fite, Stauffacher, Ostrov, and Colder (2008)	69	12.9	Self	4	F	United States	28	.28	.02
Foo (2002)	101	9.0	Parent	2	F	United States	26	.40	−.37
Foo (2002)	101	9.0	Teacher	2	F	United States	26	.40	−.08
Foster (2001); Onyskiw (1999)	12,387	7.5	Parent	2	F	Canada	—	.15	−.07
Foster (2001); Onyskiw (1999)	2,654	10.5	Self	2	F	Canada	—	.27	.02
Gleason, Jensen-Campbell, and Richardson (2004)	59	12.3	Self	4	F	United States	32	.26	−.21
Goldstein (2003)	104	12.9	Self	2	F	United States	19	.25	−.06
Grotpeter (1997)	120	8.0	Peer nominations	2	F	United States	39	.53	−.24
Hart et al. (1998)	207	5.1	Teacher	5	M	Russia	—	.04	.05
Hawley (2003)	163	4.3	Teacher	5	F	United States	45	.00	−.32
Hawley, Little, and Card (2008)	1,723	14.0	Peer nominations	4	F	Germany	19	.26	−.09
Hawley et al. (2008)	1,723	14.0	Self	4	F	Germany	19	.17	.10
Hayward and Fletcher (2003)	363	11.3	Peer nominations	4	F	Australia	—	.56	.02
Hektner and Swenson (2005)	340	11.4	Self	1	M	United States	8	.19	−.10
C. Henington, Hughes, Cavell, and Thompson (1998)	904	7.5	Peer nominations	4	F	United States	53	.42	.15
Hunt (2002)	1,008	9.0	Self	2	F	United States	36	.11	.08
Johnson (2003)	65	6.0	Peer nominations	2	M	United States	65	.61	.58
Johnson (2003)	65	6.0	Teacher	2	M	United States	65	.38	.00
Juliano, Werner, and Cassidy (2006)	67	4.3	Observation	4	F	United States	16	.39	−.09
Juliano et al. (2006)	67	4.3	Teacher	4	F	United States	16	.23	−.10
Killeya-Jones and Costanzo (2008)	254	12.2	Peer nominations	1	F	United States	43	.28	−.06
Killeya-Jones and Costanzo (2008)	254	12.2	Self	1	F	United States	43	.13	−.18
Krause (2005)	230	11.9	Peer rating	2	F	United States	—	.09	−.20

(*Continued*)

Table A1
Continued

Study	*N*	Age (years)	Reporter	Publication status[a]	Author gender	Country	% Ethnic minority	Direct (*r*)	Indirect (*r*)
Lafferty (2003)	98	12.0	Peer rating	2	F	United States	67	.52	.35
Lagerspetz et al. (1988)	167	11.5	Peer rating	4	F	Finland	—	.40	−.26
Lancelotta and Vaughn (1989)	98	9.4	Peer nominations	5	M	United States	35	.34	.40
Landau, Björkqvist, Lagerspetz, Österman, and Gideon (2002)	630	11.8	Peer ratings	4	M	Israel	—	.41	−.05
Little, Jones, Henrich, and Hawley (2003)	1,723	13.7	Self	4	M	Germany	17	.16	.10
Long (2003)	68	9.9	Teacher	2	F	United States	55	.31	−.35
Loukas, Paulos, and Robinson (2005)	745	11.8	Self	4	F	United States	21	.34	.01
McCay (2004)	440	13.6	Peer nominations	2	F	United States	6	.41	−.22
McEvoy, Estrem, Rodriguez, and Olsen (2003)	59	4.8	Observation	4	F	United States	8	.92	.64
McEvoy et al. (2003)	59	4.8	Peer nominations	4	F	United States	8	.89	.77
McEvoy et al. (2003)	59	4.8	Teacher	4	F	United States	8	1.00	.02
McNeill (2001)	415	6.8	Teacher	2	F	United States	92	.25	−.11
McNeilly-Choque, Hart, Robinson, Nelson, and Olsen (1996)	196	4.9	Observation	4	F	United States	—	.24	−.20
McNeilly-Choque et al. (1996)	241	4.9	Peer nominations	4	F	United States	—	.32	.00
McNeilly-Choque et al. (1996)	181	4.9	Teacher	4	F	United States	—	.17	−.13
Meadow (2001)	115	11.5	Peer ratings	2	F	United States	11	.43	.01
Miller, Vaillancourt, and Boyle (in press)	749	7.5	Teacher	4	F	Canada	3	.23	−.08
Mizokawa (1999)	267	13.0	Parent	2	F	United States	6	.23	−.21
Mizokawa (1999)	267	13.0	Self	2	F	United States	6	.30	−.05
Murray-Close (2006)	77	10.0	Teacher	2	F	United States	20	.30	−.25
Murray-Close, Crick, and Galotti (2006)	639	9.5	Peer nominations	4	F	United States	73	.28	−.08
Murray-Close et al. (2006)	639	9.5	Teacher	4	F	United States	73	.06	−.21
Musher-Eizenman et al. (2004)	771	10.9	Self	4	F	United States	38	.13	−.03
Nelson et al. (2005)	325	4.8	Peer nominations	4	M	United States	14	.31	.00
Nelson et al. (2005)	277	4.8	Teacher	4	M	United States	14	.23	−.17
O'Shea (2004)	111	10.3	Other (neutral peer)	2	F	United States	71	.36	.09
O'Shea (2004)	111	10.3	Other (friend)	2	F	United States	71	.38	.21
O'Shea (2004)	111	10.3	Peer nominations	2	F	United States	71	.41	−.11
O'Shea (2004)	111	10.3	Self	2	F	United States	71	.24	−.06
Österman et al. (1994)	404	8.4	Peer rating	4	F	Multiple	—	.30	.03
Österman et al. (1994)	404	8.4	Self	4	F	Multiple	—	.14	.02
Ostrov (2006)	64	3.7	Observation	4	M	United States	54	.24	.01
Ostrov and Crick (2007)	132	3.7	Observation	4	M	United States	25	.26	−.25
Ostrov and Crick (2007)	66	4.7	Teacher	4	M	United States	25	.47	.14
Ostrov and Keating (2004)	40	5.3	Observation	4	M	United States	10	.22	−.15
Ostrov, Crick, and Stauffacher (2006)	50	3.9	Observation	4	M	United States	28	.36	−.17

Table A1
Continued

Study	*N*	Age (years)	Reporter	Publication status[a]	Author gender	Country	% Ethnic minority	Direct (*r*)	Indirect (*r*)
Ostrov, Woods, Jansen, Casas, and Crick (2004)	60	4.6	Observation	4	M	United States	20	.13	−.21
Pakaslahti and Keltikangas-Järvinen (1998)	839	14.5	Peer nominations	4	F	Finland	—		
Park et al. (2005)	207	8.0	Parent	4	F	United States	7	.14	−.09
Park et al. (2005)	207	8.0	Self	4	F	United States	7	.17	.09
Park et al. (2005)	207	8.0	Teacher	4	F	United States	7	.22	−.09
Peets and Kikas (2006)	319	11.4	Peer rating	4	F	Finland	—	.66	.35
Peets and Kikas (2006)	247	11.4	Self	4	F	Finland	—	.44	.05
Peets and Kikas (2006)	335	11.4	Teacher	4	F	Finland	—	.54	.28
Pepler, Craig, and Roberts (1998)	39	9.4	Observation	4	F	Canada	57	.03	.04
Phillipsen, Deptula, and Cohen (1999)	262	8.7	Peer nominations	4	F	United States	50	.07	.00
Prinstein, Boergers, and Vernberg (2001)	566	16.4	Self	5	M	United States	78	.30	.06
Rana and Malhotra (2005)	200	16.5	Self	4	—	India	—	.34	.26
Rielly (2003)	130	11.8	Self	2	F	Canada	—	.29	.20
Russell, Hart, Robinson, and Olsen (2003)	197	4.8	Teacher	4	M	Australia	3	.41	−.09
Russell et al. (2003)	213	4.8	Teacher	4	M	United States	14	.21	−.19
Rys and Bear (1997)	266	9.5	Peer nominations	4	F	United States	36	.33	−.03
Rys and Bear (1997)	266	9.5	Teacher	4	F	United States	36	.25	−.09
Salmivalli and Kaukiainen (2004)	526	12.5	Peer rating	4	F	Finland	—	.57	.13
Salmivalli and Kaukiainen (2004)	526	12.5	Self	4	F	Finland	—	.37	.07
Salmivalli, Kaukiainen, and Lagerspetz (2000)	209	15.5	Peer nominations	4	F	Finland	—	.48	−.41
Schmidt (2004)	164	14.4	Self	2	F	United States	42	.23	.21
Schmidt (2004)	164	14.4	Other (friend)	2	F	United States	42	.12	.07
Sebanc (2003)	97	3.9	Teacher	4	F	United States	26	.23	−.21
Selah-Shayovits (2004)	921	17.3	Self	4	M	Israel	—	.37	.13
Simon (2001)	135	4.6	Peer nominations	2	F	United States	24	.61	.58
Simon (2001)	54	4.6	Teacher	2	F	United States	24	.39	.03
C. E. Smith (2004)	78	8.8	Observation	2	F	Canada	30	.46	.49
R. G. Smith (2005)	258	11.4	Parent	2	F	United States	16	.06	−.06
R. G. Smith (2005)	258	11.4	Self	2	F	United States	16	.03	.01
R. G. Smith (2005)	258	11.4	Teacher	2	F	United States	16	.18	.02
R. L. Smith, Rose, and Schwartz-Mette (2008)	607	11.1	Peer nominations	1	F	United States	13	.46	−.06
Suarez (2001)	161	5.4	Observation	2	F	United States	31	.34	.00
Suarez (2001)	161	5.4	Peer nominations	2	F	United States	31	.48	.03
Sullivan, Farrell, and Kliewer (2006)	276	14.5	Self	5	F	United States	96	.08	−.08
Tapper and Boulton (2004)	74	9.1	Observation	4	F	England	—	.18	−.19
Tapper and Boulton (2004)	74	9.1	Peer rating	4	F	England	—	.08	−.07
Tapper and Boulton (2004)	74	9.1	Self	4	F	England	—	.07	−.22
Tiet et al. (2001)	198	12.9	Parent	4	M	United States	96	.30	.00

(*Continued*)

Table A1
Continued

Study	*N*	Age (years)	Reporter	Publication status[a]	Author gender	Country	% Ethnic minority	Direct (*r*)	Indirect (*r*)
Toldos (2005)	653	15.5	Peer rating	4	—	Spain	—	.31	.11
Tomada and Schneider (1997)	314	9.2	Peer nominations	5	F	Italy	—	.46	.14
Tomada and Schneider (1997)	314	9.2	Teacher	5	F	Italy	—	.07	.06
Underwood, Scott, Galperin, Bjornstad, and Sexton (2004)	292	12.5	Observation	5	F	United States	28	.14	−.03
Vaillancourt et al. (2003)	3,089	5.9	Parent	5	F	Canada	—	.11	−.06
Verlaan (1995)	406	11.7	Peer nominations	2	F	Canada	—	.32	.02
Weiner (2002)—Study 1	461	9.0	Self	2	F	United States	—	.19	.08
Weiner (2002)—Study 2	824	11.0	Self	2	F	United States	46	.17	.11
Werner and Nixon (2005)	1,208	10.9	Self	4	F	United States	18	.18	−.05
M. Willoughby, Kupersmidt, and Bryant (2001)	362	4.2	Teacher	5	M	United States	35	.13	.08
Wolke, Woods, Bloomfield, and Karstadt (2000)	1,982	7.6	Self	5	M	England	9	.20	.08
Woods, Wolke, Nowicki, and Hall (2008)	373	9.9	Peer nominations	1	F	England	—	.38	.26
Xie, Swift, Cairns, and Cairns (2002)	475	13.4	Other (victim)	4	F	United States	100	.06	−.22
Zahn-Waxler, Park, Essex, Slattery, and Cole (2005)	54	13.4	Self	4	F	United States	18	.42	−.10
Zimmer-Gembeck, Geiger, and Crick (2005)	458	9.5	Peer nominations	4	F	United States	26	.37	.00
Zimmer-Gembeck, Hunter, and Pronk (2007)	334	11.0	Peer nominations	4	F	Australia	—	.41	−.06
					Weighted random-effects average:			.29	−.03

Note. Effect sizes for gender differences are in the metric of Pearson's correlation (*r*) for comparability to other effect sizes in this report. Gender is coded so that positive correlations represent boys being more aggressive than girls, whereas negative correlations represent girls being more aggressive than boys. M = male; F = female.

[a] Publication status was coded as: 1 = *unpublished, not dissertation*; 2 = *unpublished dissertation*; 3 = *published book chapter*; 4 = *published peer reviewed journal article*; and 5 = *published article in top-tier journal* (see text for details).

Table A2

Studies Reporting Intercorrelations Between Direct and Indirect Aggression

Study	N	Age (years)	Reporter	Publication status[a]	Author gender	Country	% Ethnic minority	r	Corrected r
Andreou (2006)	403	11.2	Peer nominations	4	F	Greece	—	.47	.56
Arnold (1998)	110	7.5	Peer nominations	2	F	United States	54	.71	.84
Berdugo-Arstark (2002)	128	8.8	Teacher	2	F	United States	0	.55	.62
Blachman (2003)	228	9.2	Other (camp counselor)	2	F	United States	47	.59	.73
Brendgen et al. (2005)	468	6.1	Peer nominations	5	F	Canada	8	.41	.56
Brendgen et al. (2005)	468	6.1	Teacher	5	F	Canada	8	.43	.50
Butovskaya et al. (2007)	212	13.7	Peer rating	4	F	Russia	—	.58	.69
Campbell (1999)	139	10.0	Teacher	2	F	United States	42	.64	.70
Carpenter and Nangle (2006)	82	4.0	Teacher	4	F	United States	6	.27	.31
Cillessen and Mayeux (2004)	607	12.0	Peer nominations	5	M	United States	34	.64	.78
Cillessen, Jiang, West, and Laszkowski (2005)	224	16.0	Peer nominations	4	M	United States	0	.61	.73
Cillessen et al. (2005)	224	16.0	Self	4	M	United States	0	.69	.80
Côté et al. (2007)	1,183	6.4	Parent	5	F	Canada	17	.34	.50
Coyne and Archer (2005)	347	12.5	Peer nominations	4	F	England	—	.54	.58
Crain (2002)	134	10.2	Peer nominations	2	F	United States	46	.87	.91
Crick (1995)	252	9.4	Peer nominations	5	F	United States	27	.66	1.00
Crick (1996)	245	9.4	Peer nominations	5	F	United States	71	.77	.87
Crick (1996)	245	9.4	Teacher	5	F	United States	71	.77	.82
Crick (1997)	1,166	10.5	Peer nominations	5	F	United States	16	.63	.69
Crick et al. (1996)	162	10.0	Peer nominations	5	F	United States	49	.16	.36
Crick et al. (1997)	65	4.5	Peer nominations	5	F	United States	27	.42	.57
Crick et al. (1997)	65	4.5	Teacher	5	F	United States	27	.75	.79
Crick and Grotpeter (1995)	491	9.4	Peer nominations	5	F	United States	40	.54	.61
Crick, Ostrov, and Werner (2006)	245	8.5	Peer nominations	5	F	United States	40	.69	.76
Crick, Ostrov, Burr, et al. (2006)	91	4.0	Observation	4	F	United States	43	.19	.22
Crick, Ostrov, Burr, et al. (2006)	91	4.0	Peer rating	4	F	United States	43	.43	.49
Crick, Ostrov, Burr, et al. (2006)	91	4.0	Teacher	4	F	United States	43	.55	.63
Cristina (2000)	87	11.1	Peer nominations	2	F	Canada	—	.65	.74
Dane (2001)	242	11.1	Peer nominations	2	M	Canada	0	.21	.45
David and Kistner (2000)	859	9.6	Peer nominations	5	F	United States	31	.82	.88
Delveaux (2003)	425	12.5	Peer nominations	2	F	Canada	0	.43	.47
Dettling, Gunnar, and Donzella (1999)	66	5.7	Teacher	4	F	United States	—	.73	.87
Estrem (2005)	100	4.2	Teacher	2	F	United States	10	.72	.80
Evans (2005)	133	12.4	Self	2	F	United States	100	.76	.92
Finch (2001)	124	9.8	Peer nominations	2	F	United States	56	.88	.92
Foo (2002)	101	9.0	Parent	2	F	United States	26	.23	.33
Foo (2002)	101	9.0	Peer nominations	2	F	United States	26	.40	.48
Foo (2002)	101	9.0	Teacher	2	F	United States	26	.41	.45
Geiger (2003)	458	9.5	Peer nominations	2	F	United States	.26	.60	.66
Gleason et al. (2004)	74	12.3	Self	4	F	United States	.32	.61	.71
Hart et al. (1998)	207	5.1	Teacher	5	M	Russia	—	.62	.69
Hawley (2003)	163	4.3	Teacher	5	F	United States	45	.57	.70
Hawley, Little, and Card (2007)	929	14.7	Peer nominations	4	F	Germany	19	.67	.79
Hayward and Fletcher (2003)	363	11.3	Peer nominations	4	F	Australia	—	.60	.72
Hektner and Swenson (2005)	340	11.4	Self	1	M	United States	8	.58	.95
C. D. Henington (1996); C. Henington et al. (1998)	904	7.5	Peer nominations	4	F	United States	53	.64	.76
Johnson (2003)	65	6.0	Peer nominations	2	M	United States	65	.90	1.00
Johnson (2003)	65	6.0	Teacher	2	M	United States	65	.39	.44

(*Continued*)

Table A2
Continued

Study	N	Age (years)	Reporter	Publication status[a]	Author gender	Country	% Ethnic minority	r	Corrected r
Kaukiainen et al. (1999)	526	12.5	Peer ratings	4	M	Finland	—	.75	.80
Killeya-Jones and Costanzo (2008)	254	12.2	Peer nominations	1	F	United States	43	.69	.82
Killeya-Jones and Costanzo (2008)	254	12.2	Self	1	F	United States	43	.34	.50
Krause (2005)	230	11.9	Peer ratings	2	F	United States	—	.08	.08
Lafferty (2003)	98	12.0	Peer ratings	2	F	United States	67	.82	.86
Landau et al. (2002)	630	11.8	Peer ratings	4	M	Israel	—	.81	.99
Leff (1995)	151	9.5	Peer nominations	2	M	United States	33	.79	.94
Little et al. (2003)	1,723	13.7	Self	4	M	Germany	17	.83	.83
Lockwood (2002)	80	10.0	Peer nominations	2	F	United States	49	.86	.97
Long (2003)	91	9.9	Teacher	2	F	United States	55	.41	.46
Loukas et al. (2005)	745	11.8	Peer nominations	4	F	United States	21	.51	.69
Macgowan et al. (2002)	171	11.8	Teacher	4	M	United States	30	.45	.60
McNeilly-Choque et al. (1996)	241	4.9	Observation	4	F	United States	—	.16	.18
McNeilly-Choque et al. (1996)	181	4.9	Peer nominations	4	F	United States	—	.50	.69
McNeilly-Choque et al. (1996)	196	4.9	Teacher	4	F	United States	—	.64	.75
Meadow (2001)	115	11.5	Peer ratings	2	F	United States	11	.57	.62
Miller (2001)	150	16.0	Peer nominations	2	F	United States	44	.53	.62
Miller et al. (in press)	749	7.5	Teacher	4	F	Canada	3	.53	.63
Murray-Close and Crick (2006)	590	9.0	Peer nominations	4	F	United States	71	.70	.77
Nelson and Crick (2002)	115	9.0	Peer nominations	3	M	United States	31	.65	.78
Nelson et al. (2005)	325	4.8	Peer nominations	4	M	United States	14	.55	.66
Nelson et al. (2005)	277	4.8	Teacher	4	M	United States	14	.62	.74
Nelson, Hart, Yang, Olsen, and Jin (2006)	215	5.1	Peer nominations	5	M	United States	—	.84	.84
O'Donnell (2002)	8	10.5	Peer nominations	2	F	United States	38	.12	.15
O'Donnell (2002)	8	10.5	Other (camp counselor)	2	F	United States	38	.19	.23
Onyskiw (1999)	11,221	7.4	Parent	2	F	Canada	—	.43	.56
Onyskiw (1999)	2,921	10.5	Self	2	F	Canada	—	.50	.69
Osantowski (2001)	123	10.5	Peer nominations	U	F	United States	46	.67	.73
O'Shea (2004)	111	10.3	Peer nominations	U	F	United States	62	.61	.67
O'Shea (2004)	111	10.3	Self	U	F	United States	62	.40	.67
O'Shea (2004)	111	10.3	Other (neutral peer)	U	F	United States	62	.74	.83
O'Shea (2004)	111	10.3	Other (friend)	U	F	United States	62	.66	.81
Ostrov (2006)	64	3.7	Observations	4	M	United States	54	.21	.24
Ostrov (2008)	139	3.8	Observation	1	M	United States	42	.16	.19
Ostrov (2008)	139	3.8	Teacher	1	M	United States	42	.60	.70
Pakaslahti and Keltikangas-Jaervinen (2000)	2,002	14.5	Peer nominations	4	F	Finland	—	.58	.73
Pakaslahti and Keltikangas-Jaervinen (2000)	2,002	14.5	Teacher	4	F	Finland	—	.57	.74
Park et al. (2005)	207	8.0	Parent	4	F	United States	7	.50	.70
Park et al. (2005)	207	8.0	Self	4	F	United States	7	.66	.85
Park et al. (2005)	207	8.0	Teacher	4	F	United States	7	.63	.75
Pepler et al. (1998)	39	9.4	Observation	4	F	Canada	57	.08	.11
Prinstein et al. (2001)	566	16.4	Self	5	M	United States	78	.52	.66
Rockhill (2000)	360	9.5	Teacher	2	F	United States	4	.64	.69
Rose, Swenson, and Waller (2004)—Study 1	607	11.1	Peer nominations	5	F	United States	13	.65	.77

Table A2
Continued

Study	*N*	Age (years)	Reporter	Publication status[a]	Author gender	Country	% Ethnic minority	*r*	Corrected *r*
Rose et al. (2004)—Study 2	1,019	11.2	Peer nominations	5	F	United States	15	.70	.83
Salmivalli et al. (2000)	209	15.5	Peer nominations	4	F	Finland	—	.68	.75
Schmidt (2004)	164	14.4	Self	2	F	United States	42	.59	.81
Schmidt (2004)	158	14.4	Other (friend)	2	F	United States	42	.44	.61
Sebanc (1999)	98	3.9	Peer nominations	4	F	United States	26	.64	.96
Sebanc (1999)	98	3.9	Teacher	4	F	United States	26	.43	.46
Simon (2001)	54	4.6	Peer nominations	2	F	United States	24	.74	1.00
Simon (2001)	135	4.6	Teacher	2	F	United States	24	.60	.66
R. G. Smith (2005)	258	13.2	Parent	2	F	United States	16	.78	.93
R. G. Smith (2005)	258	13.2	Self	2	F	United States	16	.59	.70
R. G. Smith (2005)	258	13.2	Teacher	2	F	United States	16	.55	.66
Solis (1998)	145	14.0	Self	2	F	United States	18	.59	.73
Suarez (2001)	161	5.0	Observation	2	F	United States	31	.33	.94
Suarez (2001)	161	5.0	Peer nominations	2	F	United States	31	.82	.97
Sullivan et al. (2006)	276	14.5	Self	5	F	United States	98	.52	.64
Tiet et al. (2001)	308	12.9	Parent	4	M	United States	96	.76	.97
Tomada and Schneider (1997)	314	9.2	Peer nominations	5	F	Italy	—	.76	.83
Tomada and Schneider (1997)	314	9.2	Teacher	5	F	Italy	—	.55	.68
Underwood et al. (2004)	292	12.5	Observation	5	F	United States	28	.09	.11
Vaillancourt and Hymel (2006)	585	13.5	Peer nominations	4	F	Canada	7	.57	.63
Vaillancourt et al. (2003)	3,089	7.9	Parent	5	F	Canada	—	.45	.45
Verlaan (1995)	406	11.7	Peer nominations	2	F	Canada	—	.88	.92
Weiner (2002)—Study 1	461	9.0	Self	2	F	United States	—	.33	.39
Weiner (2002)—Study 2	824	11.0	Self	2	F	United States	46	.45	.55
Werner and Nixon (2005)	1,208	10.9	Self	4	F	United States	12	.56	.78
M. Willoughby et al. (2001)	362	4.2	Teacher	5	M	United States	35	.65	.65
T. Willoughby, Chalmers, and Busseri (2004)	7,290	15.6	Self	5	F	Canada	7	.60	.76
Xie, Cairns, and Cairns (2002)	510	10.2	Self	4	F	United States	23	.00	.00
Xie, Cairns, and Cairns (2002)	510	10.2	Other (victim)	4	F	United States	23	.12	.14
Zahn-Waxler et al. (2005)	54	13.4	Self	4	F	United States	18	.48	.62
Zimmer-Gembeck et al. (2007)	334	11.0	Peer nominations	4	F	Australia	—	.50	.58
					Weighted random-effects average:			.60	.76

Note. M = male; F = female.

[a] Publication status was coded as: 1 = *unpublished, not dissertation;* 2 = *unpublished dissertation;* 3 = *published book chapter;* 4 = *published peer reviewed journal article;* and 5 = *published article in top-tier journal* (see text for details).

Table A3

Studies Reporting Associations of Direct and Indirect Aggression With Internalizing Problems

Study	*N*	Age (years)	Reporter	Publication status[a]	Author gender	Country	% Ethnic minority	Direct		Indirect		Differential *d*
								r	*sr*	*r*	*sr*	
Carpenter and Nangle (2006)	82	4.0	Teacher	4	F	United States	6	.19	.21	−.02	−.09	.37
Crick (1997)	1,166	10.5	Peer nominations	5	F	United States	16	.07	.01	.10	.06	−.06
Crick et al. (1997)	65	4.5	Peer nominations	5	F	United States	27	−.06	.04	−.17	−.16	.23
Crick et al. (1997)	65	4.5	Teacher	5	F	United States	27	.05	−.24	.25	.35	−.66
Crick, Ostrov, and Werner (2006)	234	8.0	Peer nominations	5	F	United States	40	.11	.12	.05	−.06	.20
Delveaux (2003)	425	12.5	Peer nominations	2	F	Canada	0	−.07	.05	−.25	−.24	.34
Estrem (2003)	100	4.2	Teacher	2	F	United States	11	.19	−.09	.33	.28	−.42
Evans (2005)	133	12.4	Self	2	F	United States	100	.34	−.17	.45	.33	−.58
C. Henington et al. (1998)	904	7.5	Peer nominations	4	F	United States	53	−.07	−.22	.10	.23	−.49
Long (2003)	53	9.9	Teacher	2	F	United States	55	−.13	−.21	.13	.21	−.50
Loukas et al. (2005)	745	11.8	Peer nominations	4	F	United States	21	.20	−.11	.41	.37	−.58
McCay (2004)	440	13.6	Peer nominations	2	F	United States	6	−.21	−.10	−.19	−.05	−.06
McNeill (2001)	415	6.8	Teacher	2	F	United States	92	.39	.11	.42	.19	−.09
Meadow (2001)	115	11.5	Peer ratings	2	F	United States	11	.00	.00	.00	.00	.00
Murray-Close (2006)	77	10.0	Teacher	2	F	United States	20	.25	−.04	.37	.28	−.37
Onyskiw (1999)	7,399	9.7	Parent	2	F	Canada	—	.32	.12	.40	.26	−.18
Onyskiw (1999)	2,654	9.7	Self	2	F	Canada	—	.26	.09	.29	.15	−.06
Osantowski (2001)	123	10.5	Peer nominations	2	F	United States	46	.09	.03	.10	.04	−.01
Ostrov et al. (2004)	60	4.6	Observation	4	M	United States	20	−.10	−.16	.00	.12	−.30
Phillipsen et al. (1999)	262	8.7	Peer nominations	4	F	United States	50	−.40	−.21	−.35	−.07	−.17
Prinstein et al. (2001)	253	16.4	Self	5	M	United States	78	.04	.02	.05	.03	−.01
Rockhill (2000)	172	9.5	Teacher	2	F	United States	4	.06	.01	.07	.05	−.04
Simon (2001)	135	4.6	Teacher	2	F	United States	24	.38	.34	.19	−.09	.52
Verlaan (1995)	406	11.7	Peer nominations	2	F	Canada	—	.06	−.03	.09	.07	−.10
Wolke et al. (2000)	1,639	7.6	Self	5	M	England	8	−.02	−.01	−.02	.00	−.01
Xie, Farmer, and Cairns (2003)	300	12.0	Self	4	F	United States	100	.00	.00	−.03	−.03	.05
Xie et al. (2003)	300	12.0	Other (victim)	4	F	United States	100	−.03	−.03	.00	.00	−.05
Zahn-Waxler et al. (2005)	54	13.4	Self	4	F	United States	18	.00	.00	.00	.00	.00
Zimmer-Gembeck et al. (2007)	334	11.0	Peer nominations	4	F	Australia	—	.05	.05	.02	−.01	.07
					Weighted random-effects average:			.07	−.01	.10	.08	−.11

Note. M = male; F = female.

[a] Publication status was coded as: 1 = *unpublished, not dissertation*; 2 = *unpublished dissertation*; 3 = *published book chapter*; 4 = *published peer reviewed journal article*; and 5 = *published article in top-tier journal* (see text for details).

Table A4

Studies Reporting Associations of Direct and Indirect Aggression With Externalizing Problems

								Direct		Indirect		
Study	*N*	Age (years)	Reporter	Publication status^{a}	Author gender	Country	% Ethnic minority	*r*	*sr*	*r*	*sr*	Differential *d*
Emotional dysregulation/ADHD symptoms												
Berdugo-Arstark (2002)	128	8.8	Teacher	2	F	United States	0	.51	.32	.42	.13	.24
Blachman (2003)	228	9.2	Parent	2	F	United States	47	.99	.76	.37	−.20	5.20
Blachman (2003)	228	9.2	Mixed	2	F	United States	47	.62	.92	.64	−.17	−.09
Crick (1997)	1,166	10.5	Peer nominations	5	F	United States	16	.38	.53	.00	−.36	1.14
Estrem (2003)	100	4.2	Teacher	2	F	United States	11	.34	.02	.43	.27	−.29
Little et al. (2003)	1,723	13.7	Self	4	M	Germany	17	.60	.28	.53	.07	.27
Long (2003)	53	9.9	Teacher	2	F	United States	56	.40	.39	.13	−.06	.57
McNeill (2001)	415	6.8	Teacher	2	F	United States	92	.52	.40	.34	−.08	.60
McNeilly-Choque et al. (1996)	241	4.9	Peer nominations	4	F	United States	—	.46	−.13	.81	.68	−1.51
Musher-Eizenman et al. (2004)	778	10.9	Self	4	F	United States	38	.28	.17	.23	.02	.16
Ostrov et al. (2004)	60	4.6	Observation	4	M	United States	20	.00	.00	.00	.00	.00
Rielly (2003)	130	11.8	Self	2	F	Canada	—	.70	.47	.52	−.01	.72
Russell et al. (2003)	375	4.8	Teacher	4	M	Mixed	9	.07	.01	.08	.04	−.03
Simon (2001)	135	4.6	Teacher	2	F	United States	24	.65	.52	.39	−.05	.82
Solis (1998)	145	14.0	Self	2	F	United States	18	.64	.04	.85	.55	−1.02
Tiet et al. (2001)	308	12.9	Parent	4	M	United States	96	.65	.14	.63	.03	.15
M. Willoughby et al. (2001)	362	4.2	Teacher	5	M	United States	35	.65	.53	.38	−.06	.85
Wolke et al. (2000)	1,639	7.6	Self	5	M	England	8	.24	.14	.20	.03	.12
Zalecki and Hinshaw (2004)	228	9.0	Parent	5	F	United States	47	.87	.94	.33	−.49	13.26
Zalecki and Hinshaw (2004)	228	9.0	Peer nominations	5	F	United States	47	.45	.10	.50	.25	−.18
Zalecki and Hinshaw (2004)	228	9.0	Teacher	5	F	United States	47	.83	.79	.42	−.32	2.57
Zalecki and Hinshaw (2004)	228	9.0	Other (camp staff)	5	F	United States	47	.35	−.09	.55	.43	−.67
					Weighted random-effects average:			.52	.31	.42	.06	.35
Delinquency/conduct disorder												
Blachman (2003)	228	9.2	Other (camp staff)	2	F	United States	47	.90	.47	.79	.19	.73
Costanzo, Killeya-Jones, Schmid, Mack, and Golonka (2007)	288	12.2	Peer nominations	1	M	United States	60	.17	.12	.13	−.03	.15
Costanzo et al. (2007)	288	12.2	Self	1	M	United States	60	.24	.12	.28	.18	−.07
Crick, Ostrov, and Werner (2006)	234	8.0	Peer nominations	5	F	United States	40	.49	.27	.41	.03	.29
Long (2003)	53	9.9	Teacher	2	F	United States	55	.18	.17	.07	−.02	.22
McNeill (2001)	415	6.8	Teacher	2	F	United States	92	.62	.23	.62	.24	−.01
Solis (1998)	145	14.0	Self	2	F	United States	18	.62	.15	.71	.38	−.35
Sullivan et al. (2006)	276	14.5	Self	5	F	United States	98	.69	.49	.50	.07	.63
Tiet et al. (2001)	308	12.9	Parent	4	M	United States	96	.59	−.32	.70	.49	−1.26

(*Continued*)

Table A4
Continued

Study	*N*	Age (years)	Reporter	Publication status[a]	Author gender	Country	% Ethnic minority	Direct		Indirect		Differential *d*
								r	*sr*	*r*	*sr*	
Verlaan (1995)	406	11.7	Peer nominations	2	F	Canada	—	.42	.01	.44	.16	−.16
M. Willoughby et al. (2001)	362	4.2	Teacher	5	M	United States	35	.92	.72	.57	−.04	2.03
T. Willoughby et al. (2004)	7,290	15.6	Self	5	F	Canada	7	.36	.16	.34	.10	.07
Wolke et al. (2000)	1,639	7.6	Self	5	M	England	8	.27	.24	.15	−.08	.35
Xie et al. (2003)	237	197.7	Self	4	F	United States	100	.30	.31	−.04	−.08	.55
Zahn-Waxler et al. (2005)	54	13.4	Self	4	F	United States	18	.66	.57	.35	−.07	.94
					Weighted random-effects average:			.58	.27	.45	.11	.27

Note. M = male; F = female; ADHD = attention deficit hyperactivity disorder.
[a] Publication status was coded as: 1 = *unpublished, not dissertation;* 2 = *unpublished dissertation;* 3 = *published book chapter;* 4 = *published peer reviewed journal article;* and 5 = *published article in top-tier journal* (see text for details).

Table A5
Studies Reporting Associations of Direct and Indirect Aggression With Prosocial Behavior

Study	*N*	Age (years)	Reporter	Publication status[a]	Author gender	Country	% Ethnic minority	Direct		Indirect		Differential *d*
								r	*sr*	*r*	*sr*	
Blachman (2003)	228	9.2	Parent	2	F	United States	47	−.46	−.31	−.38	−.16	−0.19
Blachman (2003)	228	9.2	Mixed	2	F	United States	47	−.63	−.49	−.41	−.10	−0.58
Carpenter and Nangle (2006)	82	4.0	Teacher	4	F	United States	6	−.48	−.57	.23	.39	−1.51
Cillessen et al. (2005)	224	16.0	Peer nominations	4	M	United States	0	−.15	−.23	.01	.17	−0.44
Cillessen et al. (2005)	224	16.0	Self	4	M	United States	0	−.10	−.16	.00	.12	−0.30
Crain (2002)	131	10.2	Peer nominations	2	F	United States	46	−.51	−.38	−.38	.19	−0.69
Crick (1996)	245	9.4	Peer nominations	5	F	United States	71	−.28	−.25	−.18	.13	−0.41
Crick (1996)	245	9.4	Teacher	5	F	United States	71	−.67	−.36	−.57	−.03	−0.46
Crick (1997)	1,166	10.5	Peer nominations	5	F	United States	16	−.40	−.27	−.29	−.02	−0.30
Crick et al. (1997)	65	4.5	Peer nominations	5	F	United States	27	−.20	−.15	−.13	−.02	−0.15
Crick et al. (1997)	65	4.5	Teacher	5	F	United States	27	−.30	−.18	−.24	−.01	−0.19
Finch (2001)	124	9.8	Peer nominations	2	F	United States	56	−.39	−.41	−.25	.28	−0.80

Table A5
Continued

								Direct		Indirect		
Study	*N*	Age (years)	Reporter	Publication status[a]	Author gender	Country	% Ethnic minority	*r*	*sr*	*r*	*sr*	Differential *d*
Geiger (2003)	458	8.0	Peer nominations	2	F	United States	26	−.41	−.27	−.31	−.05	−0.26
Hawley (2003)	163	4.3	Teacher	5	F	United States	45	−.01	−.31	.31	.44	−0.91
C. Henington et al. (1998)	904	7.5	Peer nominations	4	F	United States	53	.01	.33	−.27	−.43	0.90
Johnson (2003)	74	6.0	Peer nominations	2	M	United States	65	−.65	−.17	−.64	−.12	−0.06
Johnson (2003)	74	6.0	Teacher	2	M	United States	65	−.55	−.50	−.23	−.02	−0.68
Long (2003)	53	9.9	Teacher	2	F	United States	55	−.55	−.39	−.44	−.21	−0.26
Macgowan et al. (2002)	171	11.8	Teacher	4	M	United States	30	.25	.07	.33	.23	−0.19
McNeill (2001)	415	6.8	Teacher	2	F	United States	92	−.39	−.30	−.25	.06	−0.42
Miller (2001)	150	16.0	Peer nominations	2	F	United States	44	.00	.00	.00	.00	0.00
O'Donnell (2002)	8	10.5	Peer nominations	2	F	United States	38	−.06	.02	−.53	−.53	0.86
O'Donnell (2002)	8	10.5	Other (camp staff)	2	F	United States	38	−.48	−.46	−.20	−.13	−0.48
Onyskiw (1999)	7,399	9.7	Parent	2	F	Canada	—	−.27	−.19	−.20	−.06	−0.16
Onyskiw (1999)	2,654	9.7	Self	2	F	Canada	—	−.24	−.16	−.18	−.02	−0.16
Osantowski (2001)	123	10.5	Peer nominations	2	F	United States	46	−.56	−.17	−.60	−.28	0.14
O'Shea (2004)	111	10.3	Peer nominations	2	F	United States	71	−.47	−.44	−.21	.98	−0.72
O'Shea (2004)	111	10.3	Self	2	F	United States	71	.08	−.60	.78	.13	−9.38
O'Shea (2004)	111	10.3	Other (neutral peer)	2	F	United States	71	−.49	−.32	−.37	−.06	−0.32
O'Shea (2004)	111	10.3	Other (friend)	2	F	United States	71	−.44	−.26	−.36	−.10	−0.20
Ostrov et al. (2004)	60	4.6	Observation	4	M	United States	20	.01	.07	−.06	−.09	0.17
Pepler et al. (1998)	39	9.4	Observation	4	F	Canada	57	−.11	−.11	−.05	−.04	−0.09
Phillipsen et al. (1999)	262	8.7	Peer nominations	4	F	United States	50	−.06	−.20	.09	.21	−0.44
Rockhill (2000)	172	9.5	Teacher	2	F	United States	4	−.59	−.27	−.57	−.22	−0.07
Sebanc (1999)	98	3.9	Peer nominations	4	F	United States	26	−.10	−.62	.06	.61	−1.59
Sebanc (1999)	98	3.9	Teacher	4	F	United States	26	−.14	−.26	.20	.29	−0.68
Simon (2001)	54	4.6	Peer nominations	2	F	United States	24	−.31	−.29	−.29	.28	−0.63
Simon (2001)	135	4.6	Teacher	2	F	United States	24	−.39	−.49	−.04	.30	−0.99
Weiner (2002)	824	11.0	Self	2	F	United States	46	.00	−.06	.08	.10	−0.18
Wolke et al. (2000)	1,639	7.6	Self	5	M	England	8	−.17	−.02	−.20	−.11	0.09
Xie et al. (2003)	300	12.0	Self	4	F	United States	100	−.06	−.06	.03	.03	−0.13
Xie et al. (2003)	300	12.0	Other (victim)	4	F	United States	100	−.13	−.13	−.07	−.07	−0.09
Zimmer-Gembeck et al. (2007)	334	11.0	Peer nominations	4	F	Australia	—	−.37	−.31	−.21	.01	−0.39
					Weighted random-effects average:			−.29	−.27	−.14	.11	−0.42

Note. M = male; F = female.

[a] Publication status was coded: 1 = *unpublished, not dissertation*; 2 = *unpublished dissertation*; 3 = *published book chapter*; 4 = *published peer reviewed journal article*; and 5 = *published article in top-tier journal* (see text for details).

Table A6

Studies Reporting Associations of Direct and Indirect Aggression With Peer Relations

Study	*N*	Age (years)	Reporter	Publication status[a]	Author gender	Country	% Ethnic minority	Direct		Indirect		Differential *d*
								r	*sr*	*r*	*sr*	
Peer acceptance												
Blachman (2003)	228	9.2	Parent	2	F	United States	47	−.31	−.23	−.22	−.06	−0.21
Carpenter (2002)	75	4.1	Teacher	2	F	United States	9	−.30	−.27	−.11	−.02	−0.32
Crick and Grotpeter (1995)	491	9.4	Peer nominations	5	F	United States	40	−.17	−.10	−.15	−.06	−0.04
Crick et al. (1997)	65	4.5	Peer nominations	5	F	United States	27	.20	.13	.17	.07	0.07
Crick et al. (1997)	65	4.5	Teacher	5	F	United States	27	.12	−.10	.23	.22	−0.35
Cristina (2000)	39	9.3	Peer nominations	2	F	Canada	—	−.43	−.15	−.45	−.19	0.05
David (2001)	639	9.7	Peer nominations	5	F	United States	45	−.21	−.11	−.18	.02	−0.14
Geiger (2003)	458	8.0	Peer nominations	2	F	United States	26	−.13	−.07	−.12	−.05	−0.02
Hawley, Little, and Card (2007)	929	14.7	Peer nominations	4	F	Germany	19	−.07	.01	−.09	−.06	0.07
C. D. Henington (1996)	904	7.4	Peer nominations	4	F	United States	53	−.30	−.15	−.27	−.06	−0.10
Johnson (2003)	65	6.0	Peer nominations	2	M	United States	65	−.43	.19	−.45	−.22	0.46
Johnson (2003)	65	6.0	Teacher	2	M	United States	65	−.35	−.28	−.22	−.07	−0.26
Leff (1995)	151	9.5	Peer nominations	2	M	United States	33	−.34	−.42	−.21	.33	−0.86
Miller (2001)	150	16.0	Peer nominations	2	F	United States	44	.00	−.48	−.42	−.64	1.56
Nelson et al. (2005)	196	4.8	Peer nominations	4	M	United States	14	−.10	−.03	−.11	−.06	0.04
Nelson et al. (2005)	167	4.8	Teacher	4	M	United States	14	−.03	−.13	.09	.15	−0.31
Ostrov and Keating (2004)	42	5.3	Observation	4	M	United States	10	−.18	−.19	−.08	.09	−0.31
Ostrov et al. (2004)	60	4.6	Observation	4	M	United States	20	.24	.36	.00	−.27	0.72
Phillipsen et al. (1999)	262	8.7	Peer nominations	4	F	United States	50	.17	.11	.13	.00	0.12
Rys and Bear (1997)	266	9.5	Peer nominations	4	F	United States	36	−.22	−.18	−.14	.04	−0.25
Rys and Bear (1997)	266	9.5	Teacher	4	F	United States	36	−.22	−.15	−.16	.01	−0.18
Salmivalli et al. (2000)	209	15.5	Peer nominations	4	F	Finland	—	.07	−.11	.19	.21	−0.34
Simon (2001)	135	4.6	Teacher	2	F	United States	24	−.44	−.43	−.17	.15	−0.71
Tomada and Schneider (1997)	314	9.0	Peer nominations	5	F	Italy	—	−.35	−.26	−.25	.07	−0.37
Tomada and Schneider (1997)	314	9.0	Teacher	5	F	Italy	—	.00	.00	.00	.00	0.00
Vaillancourt (2001)	299	13.5	Peer nominations	2	F	Canada	7	.11	−.33	.57	.66	−1.46
Verlaan (1995)	406	11.7	Peer nominations	2	F	Canada	—	−.36	−.05	−.37	−.10	0.06
Zalecki and Hinshaw (2004)	228	9.0	Mixed	5	F	United States	47	−.23	−.08	−.24	−.10	0.03
					Weighted random-effects average:			−.17	−.11	−.12	.01	−0.15
Peer rejection												
Blachman (2003)	228	9.2	Parent	2	F	United States	47	.64	.42	.52	.23	0.30
Crick and Grotpeter (1995)	491	9.4	Peer nominations	5	F	United States	40	.30	.22	.20	.02	0.23
Crick et al. (1997)	65	4.5	Peer nominations	5	F	United States	27	.49	.29	.44	.20	0.11
Crick et al. (1997)	65	4.5	Teacher	5	F	United States	27	.30	.27	.17	−.11	0.43

Table A6

Continued

Study	N	Age (years)	Reporter	Publication status[a]	Author gender	Country	% Ethnic minority	Direct		Indirect		Differential d
								r	sr	r	sr	
Geiger (2003)	458	8.0	Peer nominations	2	F	United States	26	.50	.15	.55	.28	−0.17
Hawley, Little, and Card (2007)	929	14.7	Peer nominations	4	F	Germany	19	.10	−.05	.16	.14	−0.20
C. D. Henington (1996)	904	7.4	Peer nominations	4	F	United States	53	.32	.10	.34	.15	−0.06
Johnson (2003)	74	6.0	Peer nominations	2	M	United States	65	.57	.54	.53	−.50	1.62
Johnson (2003)	74	6.0	Teacher	2	M	United States	65	.54	.57	.07	−.18	1.07
Leff (1995)	151	9.5	Peer nominations	2	M	United States	33	.61	.09	.62	.12	−0.04
Miller (2001)	150	16.0	Peer nominations	2	F	United States	44	.46	.14	.56	.35	−0.28
Murray-Close and Crick (2006)	590	9.0	Peer nominations	4	F	United States	20	.53	.14	.58	.26	−0.15
Nelson et al. (2005)	180	4.8	Peer nominations	4	M	United States	14	.47	.54	.09	−.29	1.09
Nelson et al. (2005)	180	4.8	Teacher	4	M	United States	14	.33	.49	−.01	−.37	1.07
Ostrov (2008)	139	3.6	Observation	1	M	United States	42	.21	.18	.18	.14	0.04
Ostrov (2008)	139	3.6	Teacher	1	M	United States	42	.62	.36	.51	.12	0.34
Ostrov and Crick (2007)	132	4.1	Observation	4	M	United States	25	.17	.16	.05	.04	0.17
Ostrov et al. (2004)	60	4.6	Observation	4	M	United States	20	.10	.16	.00	−.12	0.30
Pakaslahti and Keltikangas-Järvinen (1998)	839	14.5	Peer nominations	4	F	Finland	—	.30	.10	.33	.15	−0.07
Phillipsen et al. (1999)	262	8.7	Peer nominations	4	F	United States	50	.01	.08	−.05	−.09	0.18
Rys and Bear (1997)	266	9.5	Peer nominations	4	F	United States	36	.47	.08	.56	.30	−0.28
Rys and Bear (1997)	266	9.5	Teacher	4	F	United States	36	.32	.10	.34	.14	−0.04
Salmivalli et al. (2000)	209	15.5	Peer nominations	4	F	Finland	—	.40	.32	.25	−.08	0.47
Tomada and Schneider (1997)	314	9.0	Peer nominations	5	F	Italy	—	.48	.40	.31	−.16	0.68
Tomada and Schneider (1997)	314	9.0	Teacher	5	F	Italy	—	.00	.00	.00	.00	0.00
Werner (2000)	881	8.5	Peer nominations	2	F	United States	37	.46	.15	.48	.19	−0.05
Werner and Crick (2004)	517	8.0	Peer nominations	4	F	United States	45	.45	.15	.47	.20	−0.06
Zalecki and Hinshaw (2004)	228	9.0	Mixed	5	F	United States	47	.60	.26	.57	.18	0.11
					Weighted random-effects average:			.39	.20	.35	.09	0.10

Note. M = male; F = female.

[a] Publication status was coded as: 1 = *unpublished, not dissertation;* 2 = *unpublished dissertation;* 3 = *published book chapter;* 4 = *published peer reviewed journal article;* and 5 = *published article in top-tier journal* (see text for details).

Child Development, September/October 2008, Volume 79, Number 5, Pages 1230 – 1236

Development of Three-Dimensional Object Completion in Infancy

Kasey C. Soska
New York University

Scott P. Johnson
University of California, Los Angeles

Three-dimensional (3D) object completion was investigated by habituating 4- and 6-month-old infants (n = 24 total) with a computer-generated wedge stimulus that pivoted 15°, providing only a limited view. Two displays, rotating 360°, were then shown: a complete, solid volume and an incomplete, hollow form composed only of the sides seen during habituation. There were no reliable preferences for either test display by 4-month-olds. At 6 months, infants showed a reliable novelty preference for the incomplete test display. Infants in a control group (n = 24) not habituated to the limited-view wedge preferred neither test display. By 6 months, infants may represent simple objects as complete in 3D space despite a limited perspective. Possible mechanisms of development of 3D object completion are discussed.

Piaget (1954) pioneered the study of infants' perceptual and cognitive abilities with his foundational work on object knowledge and spatial cognition. Since these original studies, a wealth of information on visual-cognitive abilities has accrued in such areas as object permanence (e.g., Baillargeon, 1987; Diamond, 1990), spatial reasoning (e.g., Bremner, 1978; Rieser, 1979), and perceptual completion (e.g., Johnson, Bremner, et al., 2003; Kellman & Spelke, 1983). Departures from the age and motor constraints Piaget suggested originally have been claimed, sparking polemic arguments on the "true" capacities of young infants (Spelke, 1998).

Modern rationalist views of infants' understanding of the physical world have contended that infants are more sophisticated in their representations of the visual world than Piaget surmised (Spelke, Breinlinger, Macomber, & Jacobson, 1992). Studies examining infants' understandings of object continuity and solidity, for example, have suggested that infants as young as 2.5 months may have a rudimentary understanding of object permanence (Aguiar & Baillargeon, 1999). In general, those arguing in favor of the precocious nature of infants' object perception and physical understanding maintain that developmental processes are more related to infants' sensitivity to relevant features of the objects used in the displays rather than changes in the underlying mechanisms of information-processing and object knowledge (e.g., Baillargeon, 2004).

Others have argued for a more gradual emergence of object knowledge over the 1st year after birth—with infants' understanding of the physical world built upon successively more sophisticated attentional skills, an increasing ability to integrate multiple sources of information in the visual array, and emerging manual and other exploratory skills to facilitate the acquisition of physical knowledge (e.g., Colombo, 2001; Gibson, 1988; Johnson, 2003). Work on infants' perception of causality (e.g., Cohen & Oakes, 1993), perception of object unity (e.g., Johnson, 2004), and object recognition (e.g., Kraebel & Gerhardstein, 2006) has provided evidence for infants' gradually emerging object knowledge. The present study, therefore, was motivated by this growing awareness of the importance of understanding developmental processes in providing a more accurate account of infants' physical world than is offered by rationalist approaches.

We examined infants' perception of the full, volumetric form of three-dimensional (3D) objects when provided only a limited view, asking if infants represent objects seen from a limited vantage as coherent volumes or instead as consisting solely of the surfaces that are visible from the current viewpoint. This question maps onto work investigating developmental processes of object and spatial knowledge and relates to work on perceptual completion in general in infancy.

This research was supported by the National Institutes of Health Grant R01-HD40432. Portions of this work were presented at the 2006 meeting of the Vision Sciences Society, Sarasota, FL, and the 2006 International Conference on Infant Studies, Kyoto, Japan. We gratefully acknowledge the infants and parents who participated in the studies. We also thank Juliet Davidow, Melissa Rozon, and Lauren Clepper for their help in recruiting infant participants, and we thank Holly Hamilton and Alissa Gaughan for their assistance with reliability coding.

Correspondence concerning this article should be addressed to Kasey C. Soska, Department of Psychology, New York University, 6 Washington Place, New York, NY 10003. Electronic mail may be sent to kasey.soska@nyu.edu.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7905-0002

The 3D nature of our world may pose a challenge to the immature visual system. Distant objects are often occluded by nearer objects, and recognition of the partially occluded object requires synthesis of the visible portions into a coherent percept. Objects come in and out of existence as they move or observers move around them, and the far sides of opaque objects are blocked from sight by their visible portions—the problem of self-occlusion. Despite the fragmented nature of the visual world, adult observers are adept at discriminating and recognizing complete forms through spatial and spatiotemporal occlusion (Kellman & Shipley, 1991; Marr, 1982) and through self-occlusion (Tse, 2002). Thus, our visual experience is not composed of disjoint surfaces, fleeting percepts of moving objects, and incomplete, hollow volumes; rather, our visual system "fills in" the missing parts, resulting in a richer experience than what is visible directly.

The visual system of the adult is able to accomplish perceptual completion under many circumstances without apparent effort, but the visual experience of very young infants is substantially different. Four-month-olds perceive a moving, center-occluded rod as a unified object rather than two disjoint segments (Kellman & Spelke, 1983), but newborn infants perceive this display as consisting of unconnected pieces (Slater, Johnson, Brown, & Badenoch, 1996). In these investigations, infants were habituated to a display in which a rod moved laterally behind an occluder. After habituation, infants were shown two test displays, both without the occluding box: one depicting a complete, unbroken rod and in the other, two broken rod pieces of the same size as the visible portions seen previously. Consistently longer looking at either test display is interpreted as a novelty response; therefore, a posthabituation preference for the broken rod pieces would imply unity perception during habituation. Newborns preferred the unbroken rod at test, whereas older infants preferred the broken rod parts.

Other experiments examined spatiotemporal completion: the perception of occluded trajectories of moving objects. When shown a display in which a ball moves back and forth, the center portion of its trajectory occluded by a box, infants at 6 months appeared to perceive the ball continuing behind the occluder, both in an habituation experiment (Johnson, Bremner, et al., 2003) and in an oculomotor anticipation paradigm (Johnson, Amso, & Slemmer, 2003), but younger infants provided evidence of perceiving the visible parts of the trajectory only, failing to complete it, in parallel with experiments on perception of object unity described previously. Infants' visual cognition thus undergoes substantial development during the first few months after birth toward an experience of a more holistic percept of the visual environment.

These and other experiments have characterized infants' perception of spatial and spatiotemporal unity. In contrast, little is known about how infants come to understand self-occlusion and perceive volumetric unity. Work on the development of 3D object completion and infants' understanding of self-occlusion may help elucidate how the infant's visual system overcomes the problems imposed on it by the cluttered 3D world as well as further general learning principles.

A subsidiary goal of the present work is to understand better the nature and limits of 3D form perception in young infants. Early research showed that 4-month-olds could discriminate familiar objects in novel axes of rotation from novel rotating objects (Kellman, 1984). Infants in these experiments, however, could not generalize 3D form across views within static displays and required either object or observer movement to discriminate novel and familiar objects (Kellman & Short, 1987). Infants as young as 2 months are adept at picking up 3D structure from kinetic information (Arterberry & Yonas, 2000), and by 4 months, infants are sensitive to some pictorial cues that signify 3D structure (Bhatt & Bertin, 2001; Bhatt & Waters, 1998; Shuwairi, Albert, & Johnson, 2007). Infants at 2 and 4 months will complete a 3D shape when occluded in displays depicting fully rotating 3D forms (Johnson, Cohen, Marks, & Johnson, 2003). Recent work from an operant learning paradigm provides evidence that at 3 months, infants may achieve view-invariant 3D object recognition when given the opportunity to cause the movement of various 3D forms (Kraebel & Gerhardstein, 2006).

Together, this work on perceptual completion and 3D object perception shows that young infants by 4 months show some nascent understanding that objects are continuous in space and that most objects are 3D. Infants appear to continually build upon their existing knowledge and attentional skills to visually complete more complex displays and veridically perceive more complex objects. Still, it is not known when and how infants come to perceive a complete 3D volume through self-occlusion.

In the present experiment, we showed infants computer-generated displays depicting objects rotating in depth around the vertical axis (see Figure 1)—either pivoting back and forth through 15° (*limited view*) or rotating through a full 360° (*full view*). Infants were shown a limited-view display during habituation, providing the opportunity to see only a few faces of the object. Once infants met the habituation criterion, they were shown two new full-view displays.

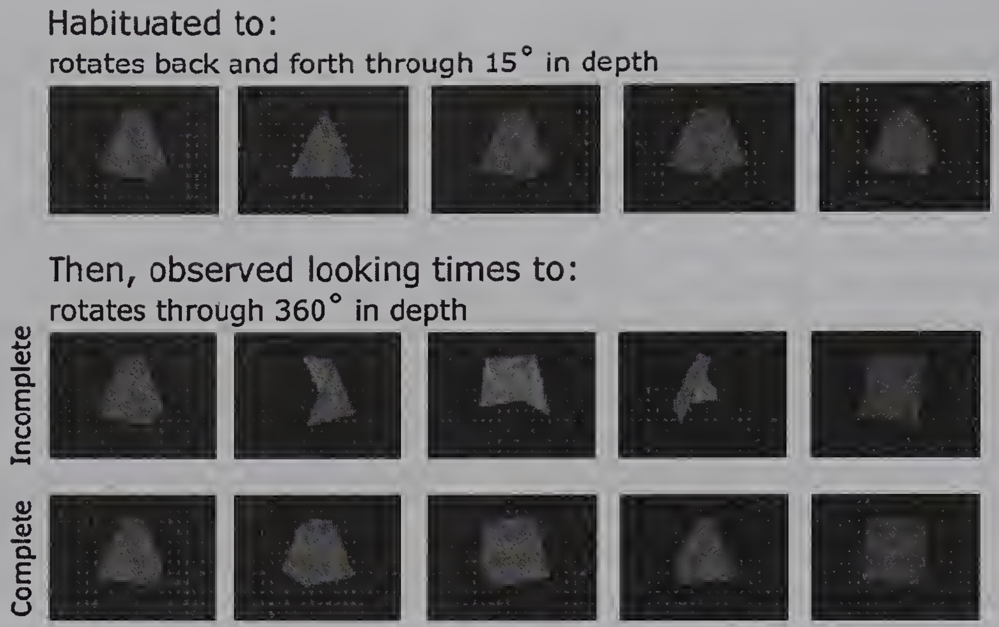

Habituated to:
rotates back and forth through 15° in depth
Then, observed looking times to:
rotates through 360° in depth
Incomplete
Complete

Figure 1. The visual stimulus set.
Note. Infants were habituated to a wedge-shaped object that rotated back and forth 15° around the vertical axis. During test, infants viewed a complete and an incomplete version of the object on alternating trials, both rotating through 360°.

One display depicted a complete solid object, and the other depicted an incomplete hollow object composed of only the surfaces seen during habituation. All the displays contained rich information for 3D form including kinetic cues (e.g., relative motion, contour deformation) and pictorial cues (e.g., shading, line junctions). We reasoned that if an infant perceived the limited-view object as complete during habituation, then during test he or she should respond to the incomplete display as novel and thus look longer at it than at the complete display. If, however, the infant showed no reliable preference for either display following habituation, then we reasoned that infant likely did not perceive the self-occluded surfaces of the limited-view object.

We tested infants at 4 and 6 months to explore developmental differences in performance. To assess whether infants in these age groups show any spontaneous preferences for incomplete or complete 3D objects, infants in a control group were shown complete and incomplete test displays without first being habituated to the limited-view object.

Method

Participants

The final sample consisted of twenty-four 4-month-olds (*M* age = 118.4 days, *SD* = 12.8) with 11 males and 13 females and twenty-four 6-month-olds (*M* age = 179.3 days, *SD* = 9.4) with 12 males and 12 females. Additional infants were observed but excluded from the analyses due to fussiness (9 infants), consistent inattention toward the screen (6 infants), parental interference (2 infants), or experimenter error (2 infants). Participants were recruited via letter and phone using a marketing database containing the names of parents with young children. The majority were from Caucasian, middle-class families in the New York City metropolitan area. All parents were fully debriefed about the experimental procedure and informed consent was obtained before beginning the session.

Apparatus and Stimuli

Infants were tested individually and seated on their parent's lap 120 cm from a 76-cm monitor in a darkened room. An experimenter was hidden behind a divider and operated a Macintosh computer that performed several functions. The computer presented the stimulus displays, stored each infant's data, calculated the habituation criterion for each infant, and changed displays after the criterion was met. The computer also recorded how long each infant looked at each display, according to the experimenter's judgments.

The displays depicted a computer-generated wedge (Figure 1). Each shape subtended a maximum visual angle of 8.5° (height) and 11.5° (width). During habituation, the object rotated 15° back and forth around the vertical axis. It took 4 s to complete a pivot in both directions and return to center. At test, following habituation, the complete object was the simplest interpretation of the limited-view object: a symmetrical mirror of the faces seen originally.

The incomplete object was composed of only the faces seen during habituation. In test displays, it took 10 s for the object to make a full rotation. The objects were covered with a brown, wood-like surface and presented against a background consisting of a 12 × 20 grid of white dots, which covered the entire monitor. Informal pilot testing revealed that all the displays were interpreted as rotating 3D objects by adult observers.

Procedure

Each trial began with the presentation of an attention-getter (a looming and contracting ball accompanied by a repetitive sequence of tones) to attract the infant's attention to the screen and center the point of gaze. Once the experimenter determined that the infant had fixated the screen, the attention-getter was replaced with the experimental stimulus and timing of each trial began. A trial ended when the infant looked away for 2 s or until 60 s of total looking had accumulated; the attention-getter was then shown again, immediately followed by a new trial. The habituation display (the limited-view object) was presented until the infant met a habituation criterion: a decline in looking time during four consecutive trials adding up to less than half the total looking time during the first four trials. Upon meeting the criterion, the infant was shown the test displays starting on the next trial. Consequently, there was a minimum of five habituation trials, and the maximum number of trials was set to be 12. The two test displays were seen in alternation three times each—for a total of six posthabituation trials. In the control task, the procedure was the same except infants were not habituated to the limited-view object display first. Instead, infants only saw the two test displays in alternation three times each. Infants were randomly assigned to one of the two test display orders (incomplete first or complete first) and to the habituation task or no-habituation control.

To assure reliability of the experimenter's judgments, an independent observer coded looking times off-line for 25% of the sample. The Pearson correlation between the experimenter's and the coder's judgments of the total looking time on each trial was .93, indicating strong agreement between the on-line and off-line coding.

Results

Data consisted of looking times toward the two displays at test. Before analysis, the data were logarithmically transformed due to excessive skew in some cells that violated the assumptions of the analysis of variance (ANOVA)—data in the figures are based on nontransformed scores. Preliminary analyses revealed there were no notable differences in looking times based on sex of infant or order of test display presentation, so these factors were not included in further analyses.

Figures 2 and 3 show looking-time results of 4- and 6-month-olds, respectively. A 2 (age: 4 or 6 months) × 2 (condition: habituation or no-habituation control) × 2 (test display: complete or incomplete object) × 3 (test trial block: first, second, or third) mixed-design ANOVA was performed—with age and condition as between-participant factors and test display and test trial block as within-participant factors. The analysis revealed a significant main effect of condition, $F(1, 44) = 16.58$, $p < .001$, partial $\eta^2 = .274$, due to greater looking overall by infants in the control condition; a significant main effect of test display,

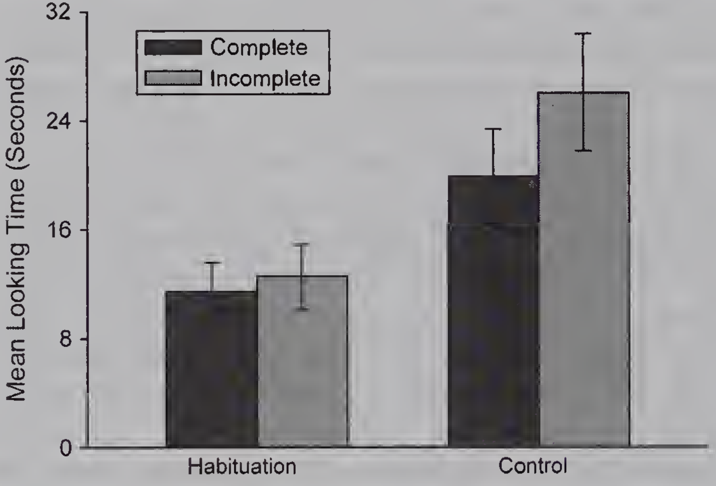


Figure 2. Looking-time results of 4-month-olds.
Note. Error bars represent standard error of the means.

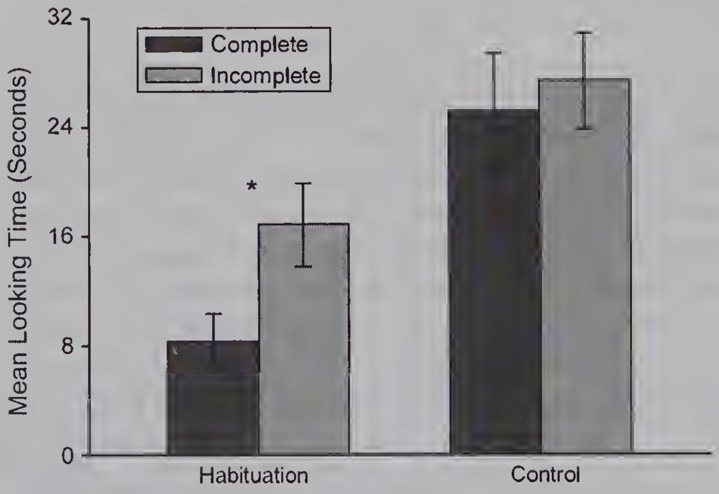


Figure 3. Looking-time results of 6-month-olds.
Note. Error bars represent standard error of the means.

$F(1, 44) = 25.11$, $p < .001$, partial $\eta^2 = .363$, due to greater looking overall at the incomplete object; and a significant main effect of test trial block, $F(2, 88) = 26.89$, $p < .001$, partial $\eta^2 = .379$, due to an overall decline in looking across test trials. This latter effect was mediated by a reliable Test Trial Block × Condition interaction, $F(2, 88) = 3.83$, $p < .05$, partial $\eta^2 = .08$, stemming from a greater decline in looking across trials by infants in the control condition.

Most important, there was a reliable three-way interaction of age, condition, and test display, $F(1, 44) = 5.11$, $p < .05$, partial $\eta^2 = .104$. Further analyses indicated that this three-way interaction was due to the presence of a significant Condition × Test Display interaction for the 6-month-old infants, $F(1, 22) = 12.00$, $p < .01$, partial $\eta^2 = .353$, but not for the 4-month-old infants, $F(1, 22) = .035$, *ns*, partial $\eta^2 = .002$. Separate pairwise comparisons in the habituation and control conditions revealed that 4-months-olds showed no reliable preference for either test display, $t(11) < 2.2$, $p > .05$ (see Figure 2). Infants at 6 months of age in the control task showed no preference for either test display, $t(11) = 0.836$, *ns*, though 6-month-old infants in the habituation condition reliably preferred the incomplete display, $t(11) = 6.16$, $p < .001$ (see Figure 3).

We conclude that infants at 4 months do not perceive completion of the self-occluded surfaces in limited-view objects. Although previous work has indicated that 4-month-olds may be sensitive to both kinetic and static cues to depth (Arterberry & Yonas, 2000; Bhatt & Bertin, 2001; Kellman, 1984; Shuwairi et al., 2007), infants at this age provide no evidence of sensitivity to the full, volumetric forms of limited-view 3D objects under the conditions we provided. Results of 6-month-olds, in contrast, provide positive evidence for 3D object completion in this age group.

Discussion

We investigated developments in 3D object completion: perception of objects as volumetric and complete in 3D space when only a limited vantage point is available. We habituated infants at 4 and 6 months to displays depicting objects rotating through 15°. During test, infants were shown alternating displays depicting objects, either complete volumes or hollow facades, rotating through 360°. If infants perceived a complete volume with self-occluded surfaces while viewing the limited-vantage object, we reasoned that the incomplete test display should have been experienced as novel and therefore recruit increased looking times relative to the complete display. The 4-month-olds preferred neither test display, providing no evidence for 3D object completion at this age, but 6-month-olds showed reliably longer looking times to the incomplete test display. These older infants provided evidence for perceptual completion of the limited-view stimulus, inferring the existence of self-occluded surfaces. Infants at both 4 and 6 months of age who were not first habituated to the limited-view object showed no reliable preferences for either display—implying that intrinsic preferences for incomplete objects can be excluded as an alternative explanation of infants' performance during the habituation tasks.

Thus, we have evidence for a developmental progression of infants' 3D object completion abilities, strengthening the view that object knowledge emerges gradually over the 1st year after birth. Between 4 and 6 months of age, infants seem to become increasingly proficient at forming percepts of complete volumes. The period around 6 months appears to represent a transitional period in the development of this visual-cognitive skill, as infants' nascent volumetric completion abilities are first becoming evident. Although it is clear that the rudiments of 3D object completion emerge around 6 months of age, further developments in this ability may relate to the complexity of the 3D forms used during testing.

These findings provide evidence regarding the time course of infants' perception of complete forms through self-occlusion to accompany the sizable literature on infants' and adults' perceptual completion abilities (e.g., Johnson, 2004; Kellman & Shipley, 1991; Tse, 2002). More than simply perceiving and registering the visible elements in a scene, perceptual completion through various types of occlusion (Johnson, Bremner, et al., 2003; Kellman & Spelke, 1983) involves the workings of sophisticated and complex mechanisms that fit together the visible parts of the world to form sensible holistic forms. 3D object completion is a perceptual inference that requires infants to recognize the unified and reversible aspects of an object in space.

Understanding the development of 3D object completion in infancy shapes a framework of how object knowledge and perception emerge over the 1st year after birth. Insights into the general processes of development can be gained, as this collective work on visual completion can help inform theories of cognitive development. For example, previous work on infants' visual completion provides strong support for a constructivist view of perceptual development (Johnson, 2003, 2004). Continually more complex forms are derived from the processing of smaller units that are assembled into a fuller percept. The experiments presented here add to this body of knowledge, by showing that 3D object completion perception may be built out of lower perceptual knowledge—kinetic and pictorial information for three dimensionality. Interestingly, infants at around 6 months of age also begin to use similarity of form to organize abstract visual patterns (Quinn, Bhatt, Brush, Grimes, & Sharpnack, 2002). Perhaps bottom-up configural processing, such as grouping by form, is developing in tandem with experiential components, supporting more sophisticated object perception such as volumetric completion.

One possibility is that infants around 6 months, who show the first evidence of 3D object completion, may also be starting to tune into the relevant features of 3D objects during normal visual experience. An object recognition system emerging during the middle of the 1st year (Colombo, 2001) may guide infants' information-processing and attentional allocation. Another possibility builds on Piaget's (1954) proposal that object knowledge is constructed in tandem with the increasing coordination of action systems. The present work on 3D object completion has discovered a perceptual ability whose onset is commensurate with the 5- to 6-month age range during which coordinated visual-manual exploration of objects is known to arise (Eppler, 1995; Rochat, 1989). As they play, infants may discover that objects have hidden surfaces, revealed via rotation, and build up volumetric completion from these explorations.

References

Aguiar, A., & Baillargeon, R. (1999). 2.5-month-old infants' reasoning about when objects should and should not be occluded. *Cognitive Psychology, 39*, 116 – 157.

Arterberry, M. E., & Yonas, A. (2000). Perception of three-dimensional shape specified by optic flow by 8-week-old infants. *Perception & Psychophysics, 62*, 550 – 556.

Baillargeon, R. (1987). Object permanence in 3.5 and 4.5-month-old infants. *Developmental Psychology, 23*, 655 – 664.

Baillargeon, R. (2004). Infants' physical world. *Current Directions in Psychological Science, 13*, 89 – 94.

Bhatt, R. S., & Bertin, E. (2001). Pictorial cues and three-dimensional information processing in early infancy. *Journal of Experimental Child Psychology, 80*, 315 – 332.

Bhatt, R. S., & Waters, S. E. (1998). Perception of three-dimensional cues in early infancy. *Journal of Experimental Child Psychology, 70*, 207 – 224.

Bremner, J. G. (1978). Egocentric versus allocentric spatial coding in nine-month-old infants: Factors influencing the choice of code. *Developmental Psychology, 14*, 346 – 355.

Cohen, L. B., & Oakes, L. M. (1993). How infants perceive a simple causal event. *Developmental Psychology, 29*, 421 – 433.

Colombo, J. (2001). The development of visual attention in infancy. *Annual Review of Psychology, 52*, 337 – 367.

Diamond, A. (1990). The development and neural basis of memory functions as indexed by the AB and delayed response tasks in human infants and adult monkeys. *Annals of the New York Academy of Sciences, 608*, 267 – 317.

Eppler, M. A. (1995). Development of manipulatory skills and the deployment of attention. *Infant Behavior and Development, 18*, 391 – 405.

Gibson, E. J. (1988). Exploratory behavior in the development of perceiving, acting, and the acquiring of knowledge. *Annual Review of Psychology, 39*, 1 – 41.

Johnson, S. P. (2003). The nature of cognitive development. *Trends in Cognitive Sciences, 7*, 102 – 104.

Johnson, S. P. (2004). Development of perceptual completion in infancy. *Psychological Science, 15*, 769 – 775.

Johnson, S. P., Amso, D., & Slemmer, J. A. (2003). Development of object concepts in infancy: Evidence for early learning in an eye tracking paradigm. *Proceedings of the National Academy of Sciences (USA), 100*, 10568 – 10573.

Johnson, S. P., Bremner, J. G., Slater, A., Mason, U., Foster, K., & Cheshire, A. (2003). Infants' perception of object trajectories. *Child Development, 74*, 94 – 108.

Johnson, S. P., Cohen, L. B., Marks, K. H., & Johnson, K. L. (2003). Young infants' perception of object unity in rotation displays. *Infancy, 4*, 285 – 295.

Kellman, P. J. (1984). Perception of three-dimensional form by human infants. *Perception & Psychophysics, 36*, 353 – 358.

Kellman, P. J., & Shipley, T. F. (1991). A theory of visual interpolation in object perception. *Cognitive Psychology, 23*, 141 – 221.

Kellman, P. J., & Short, K. R. (1987). Development of three-dimensional form perception *Journal of Experimental Psychology: Human Perception and Performance, 13*, 545 – 557.

Kellman, P. J., & Spelke, E. S. (1983). Perception of partly occluded objects in infancy. *Cognitive Psychology, 15*, 483 – 524.

Kraebel, K. S., & Gerhardstein, P. C. (2006). Three-month-old infants' object recognition across changes in viewpoint using an operant learning procedure. *Infant Behavior and Development, 29*, 11 – 23.

Marr, D. (1982). *Vision*. New York: Freeman.

Piaget, J. (1954). *The construction of reality in the child.* New York: Basic Books.

Quinn, P. C., Bhatt, R. S., Brush, D., Grimes, A., & Sharpnack, H. (2002). Development of form similarity as a Gestalt grouping principle in infancy. *Psychological Science, 13*, 320 – 328.

Rieser, J. J. (1979). Spatial orientation of six-month-old infants. *Child Development, 50*, 1078 – 1087.

Rochat, P. (1989). Object manipulation and exploration in 2- to 5-month-old infants. *Developmental Psychology, 25*, 871 – 884.

Shuwairi, S. M., Albert, M. K., & Johnson, S. P. (2007). Discrimination of possible and impossible objects in infancy. *Psychological Science, 18*, 303 – 307.

Slater, A., Johnson, S. P., Brown, E., & Badenoch, M. (1996). Newborn infants' perception of partly occluded objects. *Infant Behavior and Development*, *19*, 145–148.

Spelke, E. S. (1998). Nativism, empiricism, and the origins of knowledge. *Infant Behavior and Development*, *21*, 181–200.

Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological Review*, *99*, 605–632.

Tse, P. U. (2002). A contour propagation approach to surface filling-in and volume formation. *Psychological Review*, *109*, 91–115.

Child Development, September/October 2008, Volume 79, Number 5, Pages 1237–1256

Young Children's Recognition of the Intentionality of Teaching

Margalit Ziv and Ayelet Solomon
Tel Aviv University

Douglas Frye
University of Pennsylvania

Two studies examined the role of intention in preschoolers' understanding of teaching. Three- to 5-year-olds judged stories in which there was an intention to teach or not (teaching vs. imitation) for 4 different learning outcomes (successful, partial, failed, and unknown). They also judged 2 stories with embedded instructional intent (e.g., guided discovery learning) and several standard theory of mind tasks. There was an age-related change in the understanding of teaching. Five-year-olds distinguished teaching from imitation and recognized guided discovery learning. Understanding of imitation and false belief was related. The findings indicate that theory of mind is relevant to other means of knowledge acquisition besides perceptual access and that understanding intention could help young children to recognize instruction and identify its different forms.

Developmental psychology's recent, intense interest in children's theory of mind has provided new insights into how young children understand knowledge and the ways in which it is acquired. Preschool theory of mind can be seen as the first step in a series of changes in the understanding of knowledge (Kuhn, 2000; Montgomery, 1992; Olson & Bruner, 1996). The numerous studies that have concentrated on the basic step of false-belief understanding (see Flavell & Miller, 1998, for a review) have now shown that 5-year-olds have a different grasp of knowledge than children just 2 years younger because they realize that mistakes in representation are possible (Wimmer & Perner, 1983).

More is involved in this change in theory of mind than just the recognition of the different mental state of false belief. There is also a change in the understanding of *how* knowledge is acquired. Preschoolers gain a better understanding of perception as a means of acquiring knowledge (Pillow, 1989; Povinelli & deBlois, 1992; Taylor, 1988; Wimmer, Hogrefe, & Sodian, 1988). In the typical false-belief situation, 5-year-olds realize that it is because of a lack of perceptual access that someone will be mistaken about the location of an object or contents of a container. Three-year-olds, even though they notice the lack of perceptual access, do not seem to use the information to draw this conclusion.

New research in theory of mind has now widened the investigation of both types of knowledge and forms of knowledge acquisition young children understand. Miller (2000) integrates the findings on children's understanding of differences in many kinds of knowledge, not just false belief. Understanding of other forms of knowledge change is being studied for inference (Pillow, 1999; Sodian & Wimmer, 1987), communication and testimony (Beal, 1988; Clement, Koenig, & Harris, 2004), learning (Sobel, Li, & Corriveau, 2007), and teaching (Astington & Pelletier, 1996; Kuhn, 2000; Olson & Bruner, 1996; Strauss, Ziv, & Stein, 2002; Ziv & Frye, 2004).

Several questions are important regarding these other forms of knowledge acquisition. Accounting for young children's understanding will require finding out which forms of knowledge acquisition they grasp beyond the established "seeing leads to knowing" of perception. It will also be useful to know whether they gain an understanding of these other forms at about the same time as perception or only later. Last, because the forms do differ and may involve different mental states, it will be important to begin to determine how children identify any given one of them.

Teaching is a common way for children to acquire knowledge. It is known that even 3-year-olds refer to teaching, although not with high frequency (Bartsch, Horvath, & Estes, 2003; Sobel et al., 2007). From the standpoint of theory of mind, teaching can be defined as an intentional activity to increase the knowledge or understanding of another (Frye & Ziv, 2005). This approach extends previous analyses of the mental

---

The research was supported by a grant from the Israel Science Foundation (797/02).

Correspondence concerning this article should be addressed to Margalit Ziv, School of Education, Tel Aviv University, Ramat Aviv, Tel Aviv, Israel 69978, or to Douglas Frye, Graduate School of Education, University of Pennsylvania, 3700 Walnut Street, Philadelphia, PA 19104-6216. Electronic mail may be sent to mziv@post.tau.ac.il or to dfrye@gse.upenn.edu.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7905-0003

state aspects of teaching. Olson and Bruner (1996) state that in order to teach a person needs a theory of mind to recognize the lack of knowledge in others. Thus, awareness of a difference in knowledge is a prerequisite for teaching. Kruger and Tomasello (1996) emphasized that teaching involves intention and that it can only be adequately defined by its goal. The intent in teaching is that another person learns a new skill, acquires knowledge, or deepens his or her understanding.

The need to understand knowledge states to identify teaching has been recognized. Olson and Bruner (1996) predicted that false-belief understanding should be integral to teaching. If young children do not understand when people lack or are mistaken in their knowledge, they presumably will not be able to identify the conditions under which teaching will occur. Furthermore, they will have to make this judgment from the teacher's point of view. Teaching will not occur whenever a student misunderstands something. It is only when the misunderstanding is recognized by the one furnishing the knowledge that instruction can take place.

Ziv and Frye (2004) found empirical support for this analysis. It was established that both 3- and 5-year-olds were able to indicate that teaching would be directed toward someone who lacked knowledge about how to do something and not toward someone who knew how, and that only a knowledgeable person can assume the role of teacher, even in the instance of a child teaching an uninformed adult. Only the 5-year-olds, however, seemed able to judge the situation from the teacher's perspective and could identify teachers' beliefs about learners' knowledge. For example, only the 5-year-olds recognized that teaching would not take place when the teacher thinks a student knows how to do something, but the person really does not. The overall results showed that after 3 years, children understand that a difference in knowledge between two parties is necessary for teaching. After 5 years, they further understand that it is the teacher's belief about the knowledge difference that will actually govern teaching. This success was correlated with an understanding of false belief.

Strauss et al. (2002) presented a subset of these same tasks and found that the instructional strategies children employed during peer teaching were associated with the children's own understanding of knowledge states. Children were taught a simple board game and asked to teach and then play it with a same-aged peer. The 3-years-olds taught the peer how to play the new game by demonstrating and modeling. They also had difficulty understanding mistaken judgments of students' knowledge in the teaching stories. In contrast, 5-year-olds employed more advanced verbal teaching strategies, which included asking students about their understanding. In addition, they showed good comprehension of the theory of teaching tasks. These results are consistent with other tests of young children's teaching strategies (Ashley & Tomasello, 1998; Davis-Unger, 2008; Wood, Wood, Ainsworth, & O'Malley, 1995) and with the findings that children tend first to talk about teaching as showing and then shift to describing it as telling or explaining (Astington & Pelletier, 1996).

The relevance of intention for identifying teaching has not been empirically studied, but a theoretical basis for it exists. Kruger and Tomasello's (1996) characterization of teaching as an intentional activity builds on prior approaches to the topic (Macmillan & McClellan, 1968). Nonetheless, there are current definitions of teaching in psychology that do not include intention. For example, Caro and Hauser (1992) have argued for a functional definition. They state that teaching occurs when one party modifies his or her behavior, without direct reward, in the presence of another and that the change improves the other's learning compared to what would have occurred otherwise.

The intentional and functional definitions differ in what activities will be identified as teaching. The intentional definition also makes at least one distinction among different forms of knowledge acquisition that is not made by the functional one. Depending on which definition describes young children's concept of teaching, they should only classify certain activities as instructional. For example, the intentional definition makes it possible to separate the goal from the outcome in teaching, and therefore it still identifies the activities that occur in unsuccessful teaching as a part of teaching. According to the functional definition, the learner's behavior must change as a consequence of the teacher's behavior. Therefore, it excludes instances in which teaching was attempted but was not successful and, consequently, must identify teaching retrospectively.

Intention also establishes a distinction between teaching and imitation as distinct forms of knowledge acquisition. Imitation occurs when a novice observes a knowledgeable person do something and, as a consequence, learns how to do the same thing. This process can be distinguished from teaching. An indication of the difference is that imitation can take place even when the expert has no idea that the novice is watching. The outcome can be exactly the same in both teaching and imitation; therefore, the difference is not a functional one but instead depends on intention. In teaching, the expert presents events *in order* to change the understanding of the novice, and those choices will not be made if the expert is unaware of the novice's presence.

If children rely on intention to understand what teaching is, then developmental changes in the understanding should be apparent. Preschoolers' understanding of intention is regarded an essential component of their developing theory of mind (Astington, 1991; Malle, Moses, & Baldwin, 2001; Zelazo, Astington, & Olson, 1999). Several studies have investigated when children begin to recognize intention. The studies often concentrate on children's ability to distinguish between an intentional and nonintentional act as well as between the intention that underlies people's behavior and the outcome of that behavior (Astington, 1991; Feinfield, Lee, Flavell, Green, & Flavell, 1999; Karniol, 1978; Moses, 1993).

Astington and Lee (1991), for example, had children listen to two stories. In the first story, a girl asks her mother for some breadcrumbs, takes the crumbs outside, and throws them to the birds, who then eat them. In the second story, another girl asks her mother for a sandwich, which she takes outside to eat, but some crumbs accidentally drop behind her, and the birds eat them. The two stories have the same outcome, caused in both cases by the child's behavior, but the intention to feed the birds is only present in the first. When children were asked which girl meant to feed the birds, 3-year-olds were equally likely to choose either girl. In contrast, 5-year-olds correctly said that it was the girl who asked for the crumbs and threw them at the birds who meant to feed them.

The judgments of the 3-year-olds were interpreted as reflecting a matching strategy (Shultz & Wells, 1985). This strategy simply assumes that an outcome is intentional if it is in accord with a person's goal or existing desire and unintentional if it is not. In other words, 3-year-olds understand simple goal-directed action. They are capable of determining when a familiar action has a goal and whether the goal has been fulfilled or not (Shultz & Wells, 1985; Wellman & Woolley, 1990). However, they do not yet appreciate the crucial distinction between intention and outcome and tend to assign more weight to the outcome than to the intention when judging whether a behavior is intentional or not.

A matching strategy would not allow children to appreciate that a desirable outcome, which occurs accidentally is unintentional (Astington, 1991). Schult (2002) found that accidental outcomes were the most difficult judgment about intention for 3- and 4-year-olds to make. The 3-year-olds in particular seemed to judge intention by an act's failure or success. On the other hand, Astington and Lee's and Schult's results indicate that older preschoolers are beginning to appreciate that judging the intentionality of an act does not depend on its outcome and that desired outcomes can be caused by either intentional or nonintentional acts.

The current two studies test whether children consider intention in understanding what teaching is, and hence how theory of mind may be relevant to an additional form of knowledge acquisition besides perceptual access. The first study explores the role of intention versus outcome in children's judgment of teaching. The second study then elaborates on the relations between children's understanding of intention and their developing concept of teaching.

In the first study, new tasks were designed in which the knowledge conditions for teaching were kept constant. In all the stories, there was an experienced person who knew how to do something and an inexperienced person who did not. The tasks then varied the intention–outcome relation to explore children's understanding of the role of intention as a mediating link between the knowledge difference necessary for teaching and the outcome of the interaction. The stories differed either in intention—whether the intention to teach was present or not—or in outcome—whether learning occurred or not.

One pair of tasks examined whether the presence of the intention to teach is an important consideration for young children in judging what teaching is and if it can be considered apart from the outcome of teaching. A story in the pair described teaching that successfully results in learning. The other described a child who successfully learns a new skill by imitating an expert, when it is clear that the expert is unaware of the learner. Thus, there is no intention to teach, but learning is accomplished. If children rely on intention in their judgments, the activity should not be identified as teaching because it does not involve the expert's intention to teach. However, it may be difficult for 3-year-olds to make that distinction given they have previously been shown to be influenced heavily by positive outcomes in their judgments of intentionality.

A second comparison involves cases that are similar in their intention but differ in outcome. The stories were designed to test whether children understand that teaching is defined by the intent to bring about learning, rather than by whether it is successful in fulfilling the goal or not. One story describes a teacher who intentionally teaches but the student does not acquire any new knowledge. Another version gave no information, positive or negative, about the outcome. Lack of understanding of the crucial role of intention, as opposed to the success of the outcome, may make it difficult for 3-year-olds to accept teaching with a failed or unknown outcome as teaching.

Study 1

Method

Participants. Forty-one children from six preschools in Israel participated. There were twenty-one 3- and 4-year-olds (9 boys and 12 girls, M = 46.8 months, range = 41 – 51 months) and twenty 5- and 6-year-olds (7 boys and 13 girls, M = 68.3 months, range = 63 – 74 months). The children were proficient in Hebrew and primarily from middle-class families. None of the families were new immigrants; over 85% of the parents were born in Israel.

Design. There were seven types of stories. An example of each is given in the Appendix. Six of the seven types were related to teaching. The last type assessed false-belief understanding (Wimmer & Perner, 1983). An additional type of teaching task that assessed the understanding of instructional games was also pilot tested. On the basis of this information, a pair of these tasks was incorporated in Study 2. A description of these tasks and their results are presented there.

The teaching-related stories presented different pairings between the agent's intention and the outcome of the agent's act for another's learning. Children were asked to judge whether the agent tried to teach the other character in the story or not. The agent's act either reflected an intention to teach or not. Information regarding the outcome took one of the three forms: learning occurred, learning did not occur, or no information was given about the outcome. Using these features, there were three types of actual teaching stories and three types of imitation stories:

1. Successful teaching. The agent has the intention to teach and learning occurs.
2. Successful imitation. The agent does not have the intention to teach and learning occurs.
3. Failed teaching. The agent has the intention to teach and no learning occurs.
4. Failed imitation. The agent does not have the intention to teach and no learning occurs.
5. Unknown outcome of teaching. The agent has the intention to teach and the outcome is unknown.
6. Unknown outcome of imitation. The agent does not have the intention to teach and the outcome is unknown.

The seventh type of story was an assessment of false-belief understanding in a standard change-of-location format that asked where an uninformed character would look for a displaced object.

Procedure. Because children were asked whether the agent in each story "tried to teach," children's understanding of the phrase "try to" was tested before administering the tasks. An experimenter enacted short scenarios of dolls either trying to do something and not always completing the activity or not attempting to execute the activity at all. Children were then asked which of the characters tried to do the target activity (e.g., ride a bike). The majority of children correctly identified both the attempts to perform the activity. Three 3-year-olds erred and were excluded from the study.

Children were presented with 2 stories from each of the seven types for a total of 14 stories per child. The stories were presented across two sessions separated by about a week. One story of each type was given per session. The stories were read in Hebrew and enacted using small male and female plastic figures and a variety of objects in order to reduce the linguistic demands of the tasks. The enactment visually depicted the stories' content, especially the knowledge status or what the characters could do before and after the teaching or imitation situations. Half the stories of each type described the agent performing an academic skill (e.g., writing or counting) and the other half described a nonacademic skill (e.g., tying shoe laces or jumping on one foot). Presentation order of the academic and nonacademic skills was counterbalanced across the participants.

The characters in all the story types were described as children given that preschoolers understand that children can teach in these kinds of stories (Ziv & Frye, 2004). All the stories were counterbalanced for the characters' gender. In one of the stories of each type, the teacher was a female and the learner a male; in the other story, the roles were reversed. Each story included one to three control questions to ensure that children understood the sequence of events and the knowledge status of the story characters at the beginning and at the end of the story. If children were mistaken on the control questions, the experimenter retold the story and repeated the control questions to make certain that the child could answer them correctly. All questions—control and test—were in forced-choice format with the order of the two choices counterbalanced.

The standard change-of-location false-belief task (Wimmer & Perner, 1983) was also presented twice with different materials and characters in order to assess the children's understanding of false belief.

Each false-belief story included two control questions and one test question.

Results

Scoring. Choosing the try-to-teach answer in response to the test question was counted as correct for the successful, failed, and unknown-outcome teaching stories. The did-not-try-to-teach answer was accepted as correct for the successful, failed, and unknown-outcome imitation stories. Prediction of search in the original place was correct in the change-of-location false-belief task.

McNemar tests were employed to compare the two story versions in each task type. The stories for the teaching and imitation tasks had an academic skill and a nonacademic skill version. No difference between versions was found for any of the tasks, so responses were combined, yielding scores that ranged from 0 to 2. All participants correctly answered the control questions regarding the story characters' knowledge status at the beginning and end of the stories.

Age and task differences. The mean percentage correct responses for the two age groups for each of the seven tasks are shown in Figure 1.

The six intention – outcome tasks were analyzed together because they followed a factorial design that examined the intentional (teaching) and nonintentional (imitation) situations for the three outcomes of success, failure, and no information. Correct answers that distinguished teaching from imitation could be given to these stories strictly on the basis of the presence or absence of the intention to teach. Outcome should not have been important because the children were only asked whether the character attempted or *tried* to teach. Nonetheless, the children's responses could have been affected by whether there was a successful, failed, or unknown outcome. Gender had no effect. Therefore, to examine the two potential influences of intention and outcome and to determine whether they affected each other for the two age groups, answers to the test question were analyzed in a 2 (age) × 2 (intention) × 3 (outcome) multivariate analysis of variance (MANOVA) with follow-up analyses of variance (ANOVAs).

The MANOVA found that there was a main effect of age group, $F(1, 39) = 24.03$, $p < .001$, $\eta^2 = .38$, with the 5-year-olds giving a higher number of correct answers overall. There were also two-way interactions between age group and outcome $F(2, 38) = 4.68$, $p < .05$, $\eta^2 = .20$, as well as intention and outcome,

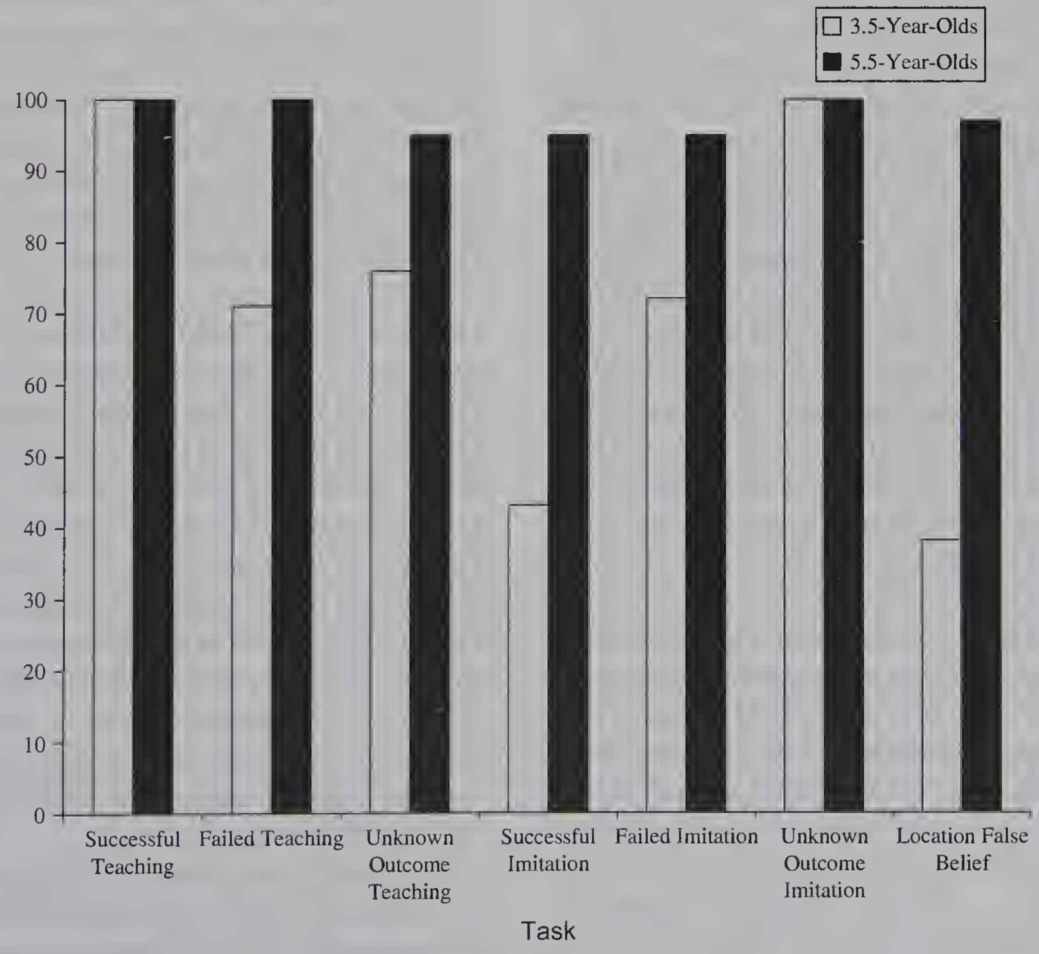


Figure 1. Mean percentage of correct responses by age for each of the seven tasks in Study 1.

$F(2, 38) = 8.87$, $p < .001$, $\eta^2 = .32$, and a three-way interaction among age group, intention, and outcome, $F(2, 38) = 7.84$, $p < .001$, $\eta^2 = .29$.

The three-way interaction was decomposed by examining the effects of the groups' judgments of teaching and imitation for each of the outcomes separately. When the successful outcome of learning was present, the 3- and 5-year-olds differed in their judgments of teaching and imitation as shown by a significant interaction between group and intention, $F(1, 39) = 8.86$, $p < .005$, $\eta^2 = .18$. The 3-year-olds did not show evidence of distinguishing between the stories on the basis of intent. When learning occurred, they said that one character tried to teach the other in both the teaching and the imitation stories. As a consequence, the 3-year-olds were more often correct on the teaching than imitation stories, $t(20) = 4.08$, $p < .001$ (two-tailed), $d = .43$. The 5-year-olds, in contrast, judged according to intention. They said that the character was trying to teach in the teaching stories and was not trying to teach in the imitation stories.

There was not a similar interaction between group and intention for either the failed or the unknown outcome stories. Overall, the 3-year-olds still showed worse performance than the 5-year-olds on stories with these outcomes, $Fs(1, 39) = 16.48$ and 11.67, $p < .001$, $\eta^2\text{s} = .30$ and $.23$, for the failed and unknown outcomes, respectively. However, the lack of Group × Intention interactions for these two outcomes indicates that the 3-year-olds were not responding at a lower level on the imitation than the teaching stories for the failed and the unknown outcomes, as they had for the successful outcome. In summary, the three-way interaction indicated that when there was a change in knowledge, 3-year-olds appeared to judge that it must be the result of teaching, and therefore mistakenly said that the character tried to teach in the imitation stories. They were much less inclined to make this mistake when there was no knowledge change or when the outcome of the story was unknown. No effect of outcome was detected for the 5-year-olds. They were able to distinguish teaching and imitation at near-ceiling levels across all three outcomes.

The results for the false-belief task were consistent with what has been found previously with these age groups (Wellman, Cross, & Watson, 2001). The 3-year-olds showed a poorer understanding of false belief because they claimed that the character would search where the hidden object was rather than where it had originally been placed. The 5-year-olds were better at determining that the character would mistakenly search at the old location, $t(39) = 5.8$, $p < .001$ (two-tailed), $d = .59$.

Comparisons to chance. The comparisons to chance largely confirmed the preceding analyses. How many children in each age group succeeded on each task was compared to chance using the binomial sign test. The 3-year-olds' responses were above chance on the successful teaching, teaching with unknown outcome, failed imitation, and imitation with unknown outcome stories. These findings add information to the comparisons among the six intention–outcome tasks. They verify that the 3-year-olds did not give accurate answers to the successful imitation story and that they were able to judge both teaching and imitation correctly when the outcome was unknown. The new information was added for the failed outcome. Although the ANOVA did not detect a Group × Intention interaction for the failed outcome stories, the 3-year-olds were only above chance on the failed imitation stories and not on the failed teaching ones. This finding indicates that the 3-year-olds were not able to say correctly that the character tried to teach when the teaching was not successful.

When considered along with the chance performance on the successful imitation task, the failed teaching results may indicate that 3-year-olds assume that teaching must be successful before it can be concluded that someone tried to teach, in addition to thinking that a successful outcome of learning is enough to say the cause was teaching.

The comparisons to chance were in agreement with the findings of the group differences for the false-belief tasks. The 3-year-olds were not above chance on false-belief indicating that they did not recognize that another would have a mistaken belief about the location of an object. The 5-year-olds' responses were above chance on all the stories.

Individual patterns of responding. The response patterns for the individual children were examined to see if they reflected the group differences. Children were given credit for passing the teaching or imitation stories if they answered at least five of the six of each. Nineteen of the 5-year-olds were successful on the teaching and imitation tasks, whereas only 7 of the 3-year-olds were, $\chi^2(1, N = 41) = 14.24$, $p < .001$. Of the 14 unsuccessful 3-year-olds, 8 passed the teaching tasks but not the imitation ones, 1 did the reverse, and the rest failed both. The individual patterns of responding verify the age differences found on the tasks and suggest that the teaching tasks as a group may have been slightly easier than the imitation ones.

Relations among the tasks. The zero-order Pearson correlations among the tasks for the 3-year-olds are shown in Table 1. Similar correlations were not computed for the 5-year-olds because their near-ceiling

Table 1

Simple Correlations Among Age and the Seven Tasks for the 3-Year-Olds in Study 1

Variable	1	2	3	4	5	6	7	8
1. Age (in months)	—	.06	−.10	−.36	−.06	.00	.12	.48*
2. Successful teaching		—	.49*	.44*	−.26	−.16	−.10	.00
3. Failed teaching			—	.62**	−.12	−.07	−.21	−.06
4. Unknown outcome teaching				—	−.03	−.16	−.35	.05
5. Successful imitation					—	.60**	.59**	.48*
6. Failed imitation						—	.72***	.08
7. Unknown outcome imitation							—	.32
8. Location false belief								—

$^{*}p < .05$ (two-tailed). $^{**}p < .01$ (two-tailed). $^{***}p < .001$ (two-tailed).

performance produced a lack of variance on many of the items.

The 3-year-olds' responses on the three teaching tasks were positively correlated with each other, $r(21) > .44$, $p < .05$, but not with anything else. Because this pattern indicates that children who answered one of the teaching tasks correctly tended to answer the other teaching ones correctly, it suggests that the subset of 3-year-olds who were successful may have been using intention to judge because the outcomes differed across the tasks. Responses on the three imitation tasks were similarly intercorrelated, $r(21) > .59$, $p < .01$. This pattern may have the similar implication that some of the 3-year-olds were using the common feature of a lack of intention to give a consistent answer across the three different outcomes. Passing the false-belief task was only related to correct judgments on the successful imitation stories, $r(21) = .48$, $p < .05$. Table 2 shows that the same correlations remained after age was partialled.

Discussion

Children's understanding of teaching shifted between 3 and 5 years, as reflected in a growing ability to distinguish teaching and imitation across different outcomes. Five-year-olds provided correct judgments across all the tasks. Three-year-olds did not distinguish between teaching and imitation for successful learning and were more influenced by outcome in general. Three-year-olds' responses were correlated within the teaching and within the imitation tasks. Their performance on the successful imitation task was related to their judgments on the standard false-belief task.

The comparison of teaching and imitation was employed in the study to test children's sensitivity to intention in judgments of teaching. Five-year-olds successfully stated that "trying to teach" occurred for teaching and not for imitation across all three outcomes—knowledge change, no knowledge change, and no information about outcome. In contrast, 3-year-olds did not separate successful teaching and successful imitation. They correctly judged successful teaching as teaching but also mistakenly regarded successful imitation as teaching. The 3-year-olds' similar judgments on the successful teaching and successful imitation stories indicate that they were using outcome to judge, and that they expected the positive outcome to have been brought about by an intentional action.

Table 2

Age-Partialed Correlations Among the Seven Tasks for the 3-Year-Olds in Study 1

Variable	1	2	3	4	5	6	7
1. Successful teaching	—	.49*	.50*	−.26	−.15	−.11	−.03
2. Failed teaching		—	.63**	−.13	−.07	−.20	.02
3. Unknown outcome teaching			—	−.06	−.16	−.34	.27
4. Successful imitation				—	.60**	.60**	.59**
5. Failed imitation					—	.72***	.09
6. Unknown outcome imitation						—	.32
7. Location false belief							—

$^{*}p < .05$ (two-tailed). $^{**}p < .01$ (two-tailed). $^{***}p < .001$ (two-tailed).

These results are similar to what Astington and Lee (1991) discovered in their study of intentional and accidental actions. They found in their example stories about birds eating breadcrumbs that young children judged that the birds could not have received the breadcrumbs by accident but had to have been given them. A similar result occurred here because the 3-year-olds judged that, when a knowledgeable person was present, the successful outcome of learning must have occurred through teaching. They did not base their judgments on intention and consequently considered imitation to be teaching. The 5-year-olds, on the other hand, realized that successful outcomes come about in different ways, through accidents or from the student's own learning, and therefore were able to differentiate the nonintentional instance of imitation (nonintentional on the knowledgeable person's part) from the intentional one of teaching.

It is interesting that the only correlation that the theory of mind task of false belief showed in the study was with the successful imitation story. Three-year-olds perform poorly on false belief because they typically make reality errors (Wimmer & Perner, 1983). The reality error in the change-of-location false-belief task is to predict that the character will look where the chocolate actually is rather than where the character originally put it. In other words, the error is to predict a successful outcome. The 3-year-olds in the current study may have shown a related response on the successful imitation task because the error on it was to assume, given a successful outcome, that an intentional act must have been responsible. This pattern is consistent with the view that 3-year-olds' understanding of theory of mind is limited by the reasoning that they can apply to intentional action (Frye, 1999). The overemphasis on a positive outcome could prevent young children from seeing how a false belief will affect someone's action and also prevent them from distinguishing between successful imitation and successful teaching.

Performance on the other tasks in the intention–outcome set was in accord with the interpretation that the 3-year-olds expected a successful outcome to be linked with the intentional act of teaching. The results of the failed teaching task fit that pattern. The 3-year-olds did not score above chance in recognizing that these stories described an attempt to teach. Compared to the successful teaching stories, the lack of a successful outcome reduced their ability to determine that the first child had tried to teach. The result demonstrates that along with assuming that a successful outcome implies teaching, 3-year-olds also seem to assume that teaching implies a successful outcome.

In contrast to young children's difficulty in judging successful imitation and failed teaching, the 3-year-olds made reasonably accurate judgments concerning situations in which no information was provided about the outcome. Their performance was at similar levels to the 5-year-olds. When the outcome was not stated, children could not use it in their judgments and could only rely on the information regarding the agent's intention. The results show that the 3-year-olds were able to use intention when it was presented this way. They performed at a level higher than chance on both the imitation and the teaching unknown-outcome stories, indicating their ability to distinguish between the two activities in situations that require consideration only of intention. Intentional teaching with no information about the output was thus judged as trying to teach, whereas imitation with no such information was not labeled teaching.

The pattern of results for all six intention–outcome tasks indicated that the more the teaching outcomes differed from success, the worse the 3-year-olds' responses were. On the other hand, the more the imitation outcomes differed from success, the better the 3-year-olds were. With regard to 3-year-olds' understanding of intention, the pattern fits a matching strategy (Astington, 1991; Schult, 2002; Shultz & Wells, 1985) and adds more information about it. The successful imitation task indicates a matching strategy because the successful outcome of learning led the 3-year-olds to the incorrect assumption that it must have been brought about by the intentional activity of teaching. The failed teaching task extends the idea of a matching strategy, but one in reverse, because when there was an unsuccessful outcome it led the 3-year-olds to the incorrect assumption that teaching had not even been attempted. However, the 3-year-olds' performance was not entirely defined by matching. When the strategy could not be used because the outcome was unknown, 3-year-olds employed the intention information in isolation to distinguish teaching and imitation. The 5-year-olds, in contrast, were not misled by outcome in any instance. They were able to use intention throughout to judge when there was an attempt to teach.

Study 2

Although the previous study found a preschool change in the use of intention to understand teaching, the timing of the improvement was difficult to determine because 3-year-olds were only compared to 5-year-olds. The current study attempted to examine the change more closely by comparing 3-, 4-, and

5-year-olds. The previous study also found that a successful outcome of knowledge change seemed to influence the 3-year-olds the most, leading them to misidentify imitation as teaching. To test this possibility further, a new task was introduced in which a partial change in knowledge was presented. If any change in skill or knowledge is sufficient to influence 3-year-olds, then the partial condition should produce a lack of discrimination between imitation and teaching again. In contrast, if it is the accomplishment of the full goal of acquiring the new skill or knowledge that is important, then the partial outcome should behave more like the failed one and permit the recognition of imitation.

Two new comparisons were included to explore how the recognition of intention is needed to understand instructional activities. In guided discovery learning (Mayer, 2004), teaching can be presented as a game in which play or exploration leads to a new skill or insight. For example, a teacher might have children play a card game in order to have them learn to recognize numbers. The intentional definition of teaching would classify this activity as teaching because of the goal involved. If there is a preschool change in understanding the intentional aspects of instruction, it could allow children to recognize the activity to be teaching rather than just play.

Intention may have similar importance in more specific instructional interactions in which knowledge is being exchanged. For instance, teachers often ask students questions, not because the teacher does not know the answer but because they want to find out if the student knows. Understanding the question's instructional significance, why it is being asked, requires understanding the underlying intent. Thus, both of these acts—instructional play and question asking—have ulterior motives that must be recognized to understand the real nature of the activity.

Finally, two new theory of mind tasks were incorporated. False belief was the only standard theory of mind assessment in the previous study. To assess the development of the understanding of intention as well, two judgment-of-intention tasks from Schult (2002) who found changes over this period were presented. The first task, intention unfulfilled/desire satisfied, involved stories in which an intentional action failed yet the desired outcome still occurred. Performance on this task might be related to understanding successful imitation because it also describes a successful outcome that had a different cause. The second type, intention fulfilled/desire unsatisfied, presented stories in which an action was carried out but, because of another event, did not result in the desired outcome. Performance on these stories might be related to understanding failed teaching because it also involves carrying out an appropriate action that does not succeed.

Method

Participants. Seventy-five children from six preschools in Israel participated. There were twenty-five 3-year-olds (10 boys and 15 girls, $M = 43.04$ months, range = 39–47 months), twenty-five 4-year-olds (13 boys and 12 girls, $M = 51.96$ months, range = 48–59 months), and twenty-five 5-year-olds (6 boys and 19 girls, $M = 66.64$ months, range = 60–75 months). The children were proficient in Hebrew and primarily from middle-class families. None of the families were new immigrants; over 85% of the parents were born in Israel.

Design. There were 11 types of stories, 6 of which were new. An example of each of the new types is given in the Appendix. The new types were as follows partial knowledge change teaching, partial knowledge change imitation, hidden intention game, hidden intention question, and two judgment-of-intention tasks from Schult (2002).

The set of teaching-related stories which presented different pairings between the agent's intention and its outcome was altered from the prior study by substituting the new partial knowledge change stories for the previous unknown outcome ones. The partial knowledge ones were added to test if it was successful knowledge change that affected young children's judgments. With the new stories, there were still six of these types, formed by the agent's act either reflecting an intention to teach or not (teaching vs. imitation) and the outcome taking the forms of successful knowledge change, partial knowledge change, or no change. The word *learning* was also removed from the teaching stories so that they closely matched the imitation ones.

The hidden intention game task tested whether teaching could be recognized despite being a part of another activity. In these stories, the agent has the intention to teach, but the instructional activity was embedded in a game. Children were asked whether the agent was using the activity to play or to teach. Another task tested whether children can understand that questions can be used for assessing knowledge rather than seeking information. Children had to say whether a question was being asked because a person who had been teaching wanted to know the answer or wanted to find out what the student knew.

Children's theory of mind understanding of intentional action outside of the realm of teaching was examined with two types of tasks from Schult (2002). Intention-unfulfilled/desire-satisfied and intention-fulfilled/desire-unsatisfied stories were presented because they exhibited age-related changes in Schult's original study and might correspond to successful imitation and failed teaching, respectively. Wording of the stories was not changed, except that only the intention test question was presented, and it was stated in terms of "try to or did not try to" in order to match the test question of the teaching stories. The other theory of mind development of false-belief understanding was again assessed using the same change-of-location format as before.

Procedure. As in the previous study, children were presented with 2 stories in each of the 11 types for a total of 22 stories per child across two sessions separated by about a week. One story of each type was given per session. For the six teaching and imitation tasks, one story of each type described the agent performing an academic skill and the other described a nonacademic skill. Presentation order of the academic and nonacademic skills was counterbalanced across the participants. The stories were read in Hebrew and enacted as in Study 1. Two versions of the intention-unfulfilled/desire-satisfied and intention-fulfilled/desire-unsatisfied tasks were presented as in Schult (2002). The standard change-of-location false-belief task (Wimmer & Perner, 1983) was again presented twice with different materials and characters.

As before, control and test questions were in forced-choice format with the order of the two choices counterbalanced for each participant.

Results

Scoring. Choosing the try-to-teach answer in response to the test question was counted as correct for the successful, partial, and failed outcome teaching stories. The did-not-try-to-teach answer was accepted as correct for the successful, partial, and failed outcome imitation stories. The really-wants-to-teach and wants-to-see-what-student-can-do replies were correct for the hidden intention game and question stories. The did-not-try-to response was correct for intention-unfulfilled/desire-satisfied stories and the did-try-to response was correct for the intention-fulfilled/desire-unsatisfied ones. Prediction of search in the original place was correct in the change-of-location false-belief task.

McNemar tests were again employed to compare the two story versions in each task type. The only difference found was for the intention-fulfilled/desire-unsatisfied stories, McNemar's $\chi^2(1, N = 75) = 8.05, p = .007$. Children were more successful on the story about shoes compared to the one about a snack. A similar difference was not found in Schult (2002). Because it is not clear if this difference is reliable or has theoretical significance, responses for this task were nonetheless combined, in line with those across the other tasks, yielding scores that ranged from 0 to 2 for each task type.

Age and task differences. The mean percentages of correct responses for the three age groups on each of the eight teaching, imitation, and hidden intention tasks are shown in Figure 2. The six intention–outcome tasks were again analyzed together because they followed a factorial design comparing the intentional (teaching) and nonintentional (imitation) situations for the three outcomes of successful, partial, and failed learning. Because gender did not have an effect, answers to the test question were analyzed in a 3 (age) × 2 (intention) × 3 (outcome) MANOVA. There was a main effect of age group, $F(2, 72) = 30.33, p < .001, \eta^2 = .46$, with success increasing with each age. There was also a two-way interaction between intention and outcome, $F(2, 71) = 57.92, p < .001, \eta^2 = .62$, and a three-way interaction among group, intention, and outcome, $F(4, 142) = 10.27, p < .001, \eta^2 = .22$.

The three-way interaction was decomposed by examining the effects of the groups' judgments of teaching and imitation for each of the outcomes separately. Significant interactions between group and intention were found for all three outcomes, $Fs(2, 72) = 15.41, 6.01$, and $5.40, p < .01, \eta^2s = .30, .14$, and $.13$, for successful, partial, and failed outcomes, respectively. To interpret these two-way interactions, the effect of group was tested for each intention separately at each of the outcomes.

With the successful outcome, there was an age difference for imitation, $F(2, 72) = 17.95, p < .005, \eta^2 = .33$, but not for teaching. Bonferroni comparisons showed better performance for the 5-year-olds than the other two groups on successful imitation. The pattern for intention was reversed for the other outcomes. There was an effect of age on partial and failed teaching, $Fs(2, 72) = 16.71$ and $16.48, p < .001, \eta^2s = .32$ and $.31$, but not on partial or failed imitation. All three age groups were different from each other for partial teaching. For failed teaching, the 5-year-olds performed better than the other two groups.

In sum, the pattern for the three-way interaction repeated and expanded what was found in Study 1. When there was a successful outcome, judging imitation was difficult for the 3- and 4-year-olds because they misidentified successful imitation to be teaching. In contrast, when there was a failed outcome, judging teaching was a problem for the 3- and 4-year-olds

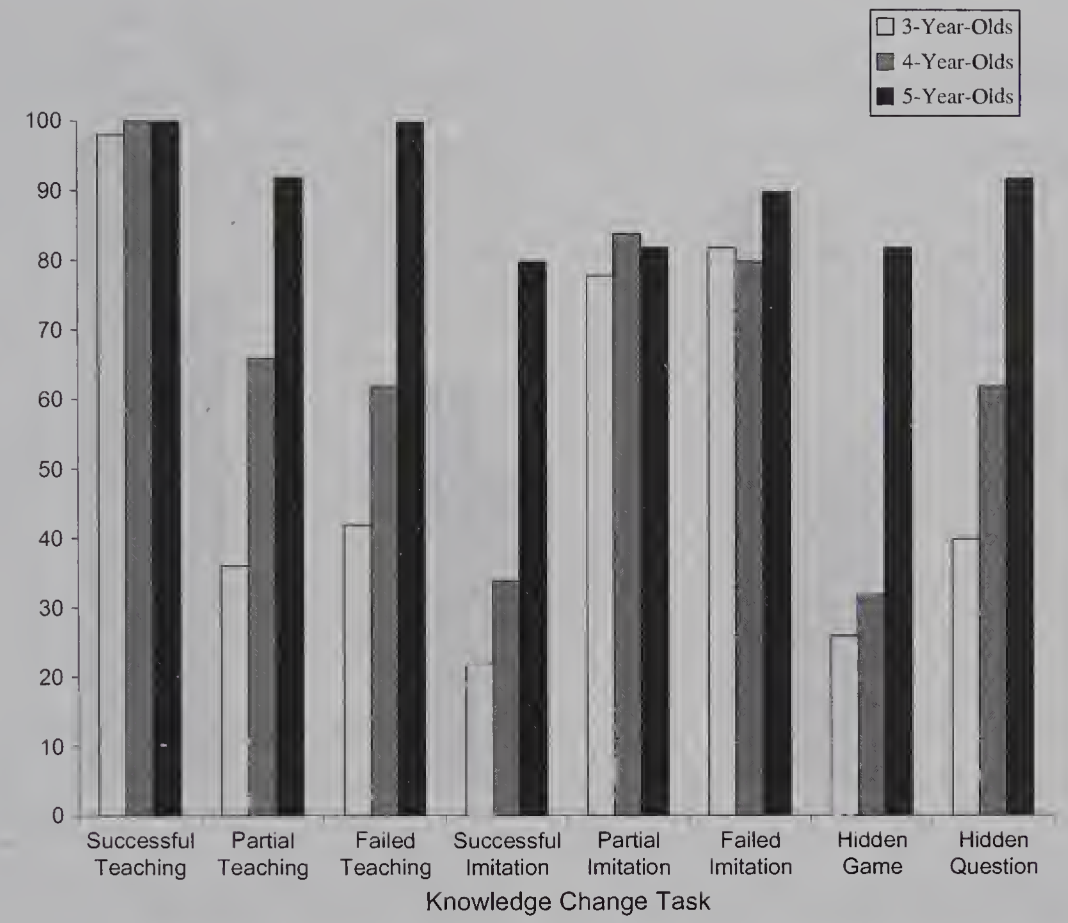


Figure 2. Mean percentage of correct responses by age for each of the eight knowledge change tasks in Study 2.

because they tended to say there had not been an attempt to teach. The new finding was that the pattern for the partial outcome resembled the failed one. The two younger groups realized that there had been no attempt to teach with partial imitation, but they also incorrectly said that there had been no attempt to teach when teaching resulted in a partial outcome.

The hidden intention game and question tasks were analyzed in a 3 (age) × 2 (task) MANOVA. Main effects were found for age, $F(2, 72) = 30.43$, $p < .001$, $\eta^2 = .46$, and task, $F(1, 72) = 9.58$, $p < .01$, $\eta^2 = .12$. Bonferroni comparisons showed that the 5-year-olds performed better than both the 3- and 4-year-olds. In addition, response levels on the question stories were higher than on the game stories. Overall, children were better at recognizing that when a teacher asks a question in these stories, the intent is to see whether the learner knows the answer than they were at understanding that when teacher plays a game with the learner the intent is to teach.

The mean percentages of correct responses for the three age groups on the false-belief, intention-unfulfilled/desire-satisfied, and intention-fulfilled/desire-unsatisfied tasks are shown in Figure 3. These responses were analyzed together because they examined different standard aspects of theory of mind. A 3 (age) × 3 (task) MANOVA found main effects for age, $F(2, 72) = 47.23$, $p < .001$, $\eta^2 = .57$, and task, $F(2, 71) = 11.78$, $p < .001$, $\eta^2 = .25$. Bonferroni comparisons revealed each of the groups improved with age, and the intention-fulfilled/desire-unsatisfied task was easier than the other two.

Comparisons to chance. For the teaching and imitation tasks, the comparisons to chance largely confirmed the preceding analyses and showed similarities between 3- and 4-year-olds' performance. Both age groups were above chance on successful teaching, partial imitation, and failed imitation. In addition, the 3-year-olds were below chance on successful imitation. These findings verify that the 3- and 4-year-olds did not give accurate answers to the successful imitation story and that they were able to judge imitation, but not teaching, correctly when the outcome was partial or failed learning. The pattern strengthens the findings of Study 1 in indicating that 3- and 4-year-olds assume that teaching must be successful before it can be concluded that someone tried to teach, in addition to thinking that a successful outcome is enough to say the cause was teaching. The 5-year-olds' responses were above chance on all six of these tasks, showing that they were able to distinguish teaching and imitation on the basis of intention regardless of outcome.

The comparisons to chance for the two hidden intention tasks followed the MANOVA results. The

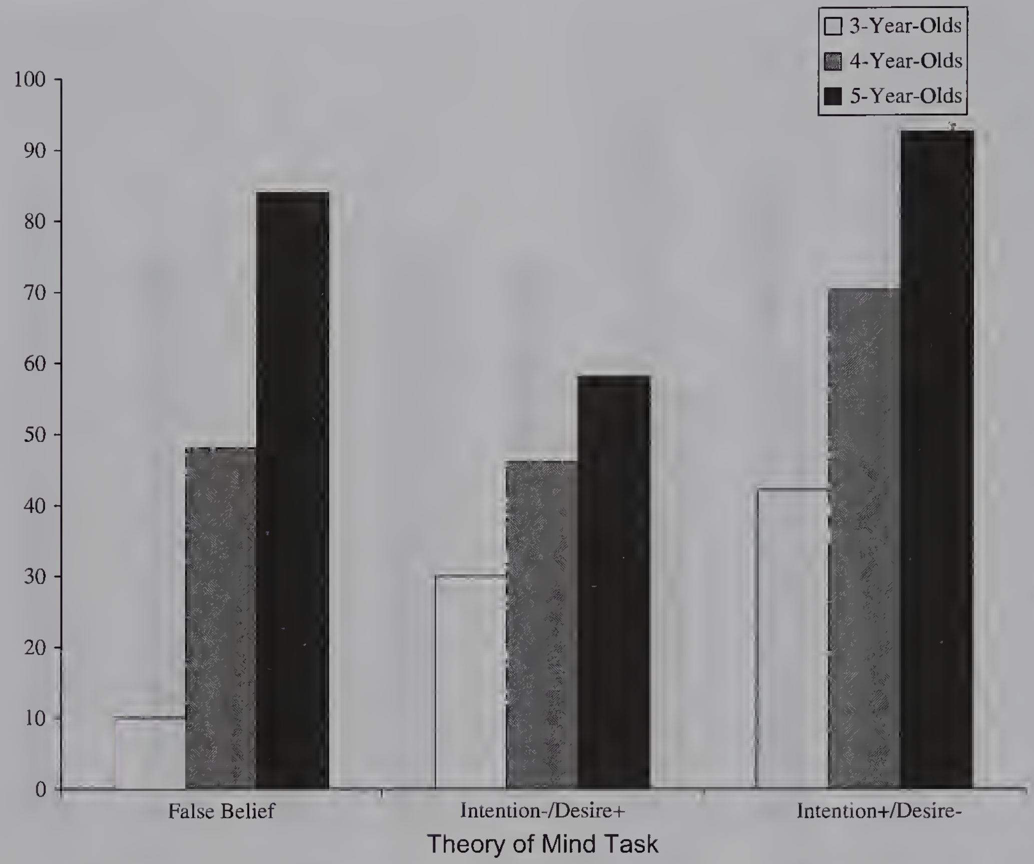


Figure 3. Mean percentage of correct responses by age for each of the three theory of mind tasks in Study 2.

3- and 4-year-olds' responses were not different from chance on either task, indicating that they did not recognize the instructional intention of playing the game or asking questions. The 5-year-olds' performance was above chance for both.

On the theory of mind tasks, the 3-year-olds' performance was below chance on false belief. Otherwise, the 3- and 4-year-olds' responses were not different from chance for either false belief or the two intention tasks. The 5-year-olds gave correct answers at above chance rates for false belief and the intention-fulfilled/desire-unsatisfied tasks, but not for the intention-unfulfilled/desire-satisfied one. Thus, even 5-year-olds had difficulty in recognizing that the satisfactory outcome in these stories was not a result of an intentional action.

Individual patterns of responding. The individual response patterns for the teaching and imitation stories were scored as in the previous study, with a minimum of five of the six stories correct needed to pass each. With the new hidden intention tasks, at least three of the four game and question stories had to be correct to pass these tasks. Using these criteria, there was one 3-year-old, three 4-year-olds, and seventeen 5-year-olds who passed the teaching, imitation, and hidden intention tasks, revealing a strong difference in the number of children who succeeded with age, $\chi^2(2, N = 75) = 30.16, p < .001$.

Of the children who did not pass all the tasks, the pattern was clear for the 5-year-olds. Of the 7 children who did not pass all the tasks, 3 passed the teaching and imitation tasks but not the hidden intention ones and 4 passed only the teaching and hidden intention ones. These were the same categories as in Study 1. The most frequent patterns overlapped for the 4-year-olds: Of the 21 children who did not pass all the tasks, 4 passed the teaching and imitation, 5 passed teaching and hidden intention, 4 passed the teaching alone, 2 passed imitation alone, and 6 did not pass any of the three. There were three patterns that were common for the 3-year-olds. Of the 24 children who did not pass all the three types, 6 only succeeded on imitation, 5 only on teaching, and 13 did not pass any of them. The individual patterns of responding show again that it was only the 5-year-olds who appeared to succeed across all the knowledge change tasks.

Relations among the tasks. The zero-order Pearson correlations and age-partialed correlations among the tasks for the 3-, 4-, and 5-year-olds are shown in Tables 3 and 4. Many of the correlations remained even after age had been partialed. The age-partialed correlations revealed separate patterns for teaching and imitation. Successful teaching was only negatively correlated with successful imitation. However, partial and failed teaching was related to each other. The imitation tasks also shared similar responding.

Table 3
Simple Correlations Among Age and the 11 Tasks in Study 2

Variable	1	2	3	4	5	6	7	8	9	10	11	12
1. Age (in months)	—	.08	.58***	.52***	.61***	.11	.12	.58***	.56***	.60***	.34**	.49***
2. Successful teaching		—	.04	−.08	−.15	−.06	−.06	.13	−.10	.12	−.17	.06
3. Partial teaching			—	.65**	.34**	−.14	−.24*	.39***	.43***	.48***	.05	.52***
4. Failed teaching				—	.35**	−.32**	−.16	.29*	.58***	.32**	.16	.64***
5. Successful imitation					—	.38**	.32**	.35**	.48***	.58***	.30*	.31**
6. Partial imitation						—	.59***	.05	.05	.12	.19	−.10
7. Failed imitation							—	.06	.16	.07	.22	−.04
8. Hidden intention game								—	.29*	.38**	.09	.33**
9. Hidden intention question									—	.44***	.31**	.38**
10. Location false belief										—	.07	.44***
11. Intention−/Desire+											—	.09
12. Intention+/Desire−												—

*$p < .05$ (two-tailed). **$p < .01$ (two-tailed). ***$p < .001$ (two-tailed).

All three were related to each other. In addition, the partial and failed teaching tasks were negatively related to the partial and failed imitation ones. These patterns are similar to Study 1 in showing that teaching and imitation tended to form two distinct sets of tasks and here responding was negatively related across the sets.

The only correlation for the hidden intention tasks was that the question task was related to failed teaching. Among the standard theory of mind tasks, false belief was again only related to successful imitation. This outcome replicates what was found in Study 1 and adds substantial support for this specific relation. With the intention tasks, intention fulfilled/desire unsatisfied was related to partial and failed teaching. This relation is plausible because in all three tasks, the actor carries out the action but does not succeed in bringing about the outcome.

Discussion

The current results support the age difference in the understanding of teaching found in the previous study. The 3-year-olds again appeared to incorporate outcome in their conception of teaching. When there was a successful outcome, they identified imitation as teaching, despite the lack of intention on the model's part. When there was no learning in the failed teaching story, they judged that there had not been an attempt to teach. The three age groups allowed for a better assessment of the timing of the change in the understanding of teaching. In general, the results for the 4-year-olds were closer to those of the 3-year-olds rather than the 5-year-olds, putting the change between 4 and 5 years. The new partial teaching and imitation situations showed that the young children only counted full learning as success. When just

Table 4
Age-Partialed Correlations Among the 11 Tasks in Study 2

Variable	1	2	3	4	5	6	7	8	9	10	11
1. Successful teaching	—	.00	−.15	−.25*	−.07	−.07	.10	−.17	.10	−.21	.02
2. Partial teaching		—	.50***	−.18	−.25*	−.38**	.09	.15	.21	−.19	.33**
3. Failed teaching			—	.05	−.45***	−.26*	−.02	.42***	.01	−.02	.52***
4. Successful imitation				—	.39**	.31**	.00	.21	.33**	.11	.02
5. Partial imitation					—	.58***	−.02	−.01	.07	.16	−.18
6. Failed imitation						—	.00	.11	−.01	.19	−.11
7. Hidden intention game							—	−.05	.05	−.13	.06
8. Hidden intention question								—	.15	.15	.15
9. Location false belief									—	−.18	.21
10. Intention−/Desire+										—	−.09
11. Intention+/Desire−											—

*$p < .05$ (two-tailed). **$p < .01$ (two-tailed). ***$p < .001$ (two-tailed).

several letters of a word were learned or a face partially drawn, the 3-year-olds' responses corresponded to those of the failed stories in stating that there had not been an attempt to teach. This finding indicates that any change in knowledge or skill was not enough to affirm an attempt to teach. The outcome appeared to have to fulfill the goal of the teaching before the young children acknowledged that there was an attempt to teach.

The other tasks exhibited similar age patterns to the teaching–imitation ones. The 5-year-olds performed better than the other two groups on the hidden intention tasks about the real point of the game or the teacher's question. Performance overall was slightly better on the question task, suggesting that perhaps children are more accustomed to interpreting the hidden intent of teacher's questions. The age differences for the theory of mind tasks were also similar. Only the 5-year-olds succeeded on false belief. Success on the judgment-of-intention tasks also showed a change over this age period, as Schult (2002) found, although the intention-unfulfilled/desire-satisfied stories were difficult even for the 5-year-olds. Similar to the teaching and imitation stories, these tasks can be viewed as presenting a disparity between intention and outcome, and 3-year-olds seem to display their characteristic pattern of responding. In the intention-fulfilled/desire-unsatisfied task, in which an intentional action was not successful, the 3-year-olds incorrectly judged that the person had not tried to bring about the outcome. In the intention-unfulfilled/desire-satisfied one, in which the successful outcome did not depend on the person's action, 3-year-olds, and many of the older children, incorrectly said that the person tried to produce the outcome. Both of these tasks again show the influence of a successful outcome on young children's judgments.

The age-partialed correlations reaffirm the links found before. Responding was related on the partial and failed teaching tasks and among the imitation tasks, although it was dissimilar across these two sets. The specific relation between understanding of false belief and successful imitation was again found. The replication of this correlation with a new sample and wider age range suggests that it is a reliable finding. It was proposed before that the two judgments might be connected because they could both require the understanding that in some instances success cannot be the sole consideration. In other words, young children who fail these tasks mistakenly make the simplest assumptions that searching will produce the hidden object and learning is brought about by teaching. By this logic, the intention-unfulfilled/desire-satisfied task should have also been related, although the overall difficulty of this task may indicate that there is more involved in it. A relation was found between the intention-fulfilled/desired-unsatisfied task and the partial and failed teaching ones. This relation is interpretable in intention–outcome terms in that all three represent instances in which the appropriate action was accomplished but a successful outcome did not follow.

General Discussion

The overall results of the two studies provide evidence of a change in preschoolers' understanding of the intentional aspect of instruction. Three- and 4-year-olds frequently judged by outcome. When there was successful learning, they judged there had been an attempt to teach, even in the case of imitation. When teaching failed, they did not say that there had been an attempt to teach. It was only when the outcome was unknown in Study 1 that they seemed to be able to use intention to identify teaching and imitation. The 5-year-olds' sensitivity to intention allowed them to distinguish teaching and imitation, recognize teaching even when it was not successful, and see the hidden instructional intent of games and questions despite their surface appearance. These findings show that recognition of intention is crucial for comprehending the common knowledge acquisition process represented by teaching and that children's understanding of teaching changes significantly over the preschool period.

The change makes a difference for the characteristics of the children's concept of teaching. Except for their performance when an outcome was unknown, the 3-year-olds had close to a functional definition of teaching. Because the functional definition relies on a successful outcome to designate teaching, it does not separate successful teaching from successful imitation or identify failed teaching as an attempt to teach. The 5-year-olds' use of intention gave them a different conceptual basis for interpreting teaching. Because they were able to consider intention, they could distinguish successful teaching and successful imitation on the basis of whether the actor was trying to change the knowledge state of the observer or not. Similarly, they were able to continue to identify failed teaching as an attempt to teach because they understood what the actor was trying to do, despite not succeeding.

Understanding an underlying intention also makes it possible to recognize instructional characteristics that are not obvious. The older preschoolers began to see that a game in the classroom could actually be an instance of instruction (i.e., guided discovery learning) because it fit with what the teacher was trying to accomplish. Seeing the intention allows a new insight into activities, such that understanding what the activity is used for rather than what it is becomes more important. Understanding the point of questions showed a similar change. There are several ways preschoolers could perceive adult questions (Donaldson, 1978). The current findings suggest that young preschoolers still see them as simple requests for information, but that 5-year-olds can recognize when they are being used to assess the child's knowledge.

Children's growing understanding of games and questions that can be embedded in teaching suggests that they gradually come to understand the *process* of teaching and the means teachers use to accomplish their goals. Recognizing the meaning of questions for teaching at the age of 5 years is especially important because it shows an awareness of the teacher's focus on the change in learners' knowledge and the need to continuously assess and respond to learners' understanding. Whereas recognition of teachers' intention—to enhance learner's knowledge or understanding based on identifying a knowledge gap between learner and teacher (Ziv & Frye, 2004)—give the necessary conditions for initiating teaching and for identifying teaching situations, understanding the meaning of teaching activities reflects an appreciation of the dynamics of teaching.

Fully understanding the dynamic aspects of selecting appropriate instructional tactics and modifying them on the basis of learners' responses to an instruction may require second order theory of mind reasoning (Wood et al., 1995). For instance, higher order, recursive reasoning is important for recognizing learners' failure to understand what a teacher tried to teach (e.g., teacher thought that X would understand that she meant him to know or do) and modification of subsequent instruction in light of this recognition. Five-year-olds' understanding of teacher's use of questions may indicate an initial appreciation of teachers' responses to learners' needs. An increased understanding would be expected between ages 5 and 7, as children improve in their understanding of the recursive nature of mental states (Perner, 1988; Perner & Wimmer, 1985).

The current results suggest that teaching should be considered among other common forms of knowledge acquisition—perceptual access, inference, communication and testimony, and learning—studied within theory of mind. It will be important to determine how children understand each of these different interactions. The present study gives an example of how the understanding of two of the forms are related in that recognizing the presence or absence of intention allowed teaching and imitation to be discriminated. Examining the mental states involved can also distinguish these forms from other theory of mind effects. For instance, teaching is different from deception, which has long been studied in theory of mind (Wimmer & Perner, 1983), not because one is intentional and the other is not, but because the intent in one is to increase knowledge and the intent in the other is to distort it. It may be that all these forms of knowledge change belong to a related family of effects that could be better understood by exploring the differences and similarities among them.

Research on children's understanding of the relation between teaching and learning seems to be the most natural next step. A central question concerns children's conceptions of the similarities and differences between the two activities. Using intention to analyze teaching draws an important distinction between teaching and learning. According to the current approach, teaching must be intentional, but the same is not true for learning. Although much of our learning is intentional, for learning to occur there merely has to be a change in the person's knowledge or understanding, regardless of the experience that caused the change (Frye & Ziv, 2005). Exploring children's understanding of the correspondence between teaching and learning will portray a richer picture of their developing understanding of knowledge acquisition processes.

The findings concerning preschoolers' understanding of teaching may also open new questions about how young children perceive the actual teaching and learning situations that they encounter everyday in preschool and perhaps at home. For instance, because 3-year-olds strongly rely on outcome, when successful learning occurs in the presence of an adult it may be perceived as teaching, even though there was no act or intention on the adult's part. Thus, 3-year-olds could perceive spontaneous or imitative learning that occurs at school and home as being caused by the adult, rather than by the learner's practice or intentional attempts to master a new skill. Certainly, when young children are asked how they learned to do something so well (e.g., put together a puzzle), they often say that their teacher, parents, or siblings taught them.

The high percentage of 3-year-olds who do not acknowledge failed teaching as an attempt to teach further emphasizes young children's strong reliance on a successful outcome in conceptualizing teaching. Their assumption seems to be that teaching always brings about learning, otherwise it is not teaching.

This view raises an interesting question of when no learning occurs, and yet "teaching" acts may have been observed by young children, how do they interpret them? When a teacher shows a child how to assemble a new toy or how to play a new game and the child does not understand, is the teacher's activity regarded as some kind of a routine communication or a common attempt to play? If young children do not appreciate what unsuccessful teaching is, it strongly suggests that they will not see the point of what the teacher was attempting to do.

Although young children's understanding of teaching is restricted compared to the understanding of older children and adults, it provides the conceptual foundations for the shift in understanding that occurs between the ages 3 and 5 years and for understanding teaching situations that are likely to be common in preschools. The 3-year-olds in the current study showed an understanding of the causal relation between teaching and learning when the outcome was successful. Furthermore, they correctly identified situations that depicted an attempt to teach when no information was given about the outcome. Children may often observe a teacher teaching a child and not know the outcome. Three-year-olds may assume a successful outcome in these situations and therefore be able to account for them as teaching, but doing so could also help them to see the teachers' acts of demonstrating, showing, and explaining as being relevant to teaching. In that case, the experiences would allow young children to learn about aspects of teaching in more instances than just when it results in obvious learning.

An additional important question of the current findings concerns the effects that understanding teaching as a form of knowledge acquisition might have for the child's own learning. Within theory of mind research, it has been shown that understanding perceptual access grants children an advantage in choosing what information to accept (Povinelli & deBlois, 1992). In fact, there is evidence that preschoolers make generalized judgments about whether a person's knowledge is trustworthy (Koenig & Harris, 2005). It has been established as well that improvements in theory of mind are related to observable changes in children's own ability to teach (Davis-Unger, 2008; Strauss et al., 2002). It is also possible that changes in the understanding of teaching make a difference for the child's ability to learn from it. There are already some positive indications of this possibility because older preschoolers learn more when they are told that they are going to be taught (Jeffrey, 2003; Taylor, Esbensen, & Bennett, 1994) and when teachers appear certain about their own knowledge (Sabbagh & Baldwin, 2001). In exploring this issue, it will also be important to consider individual differences in children's understanding of teaching and how they benefit from different learning situations.

The changes in children's understanding of teaching that are evident by 5 years may have important educational implications for formal instruction. Because 5-year-olds appear to realize that teaching is intentional, they can distinguish teaching from imitation, identify teaching despite a negative outcome, and recognize it even when it is embedded in a different activity. The first educational implication of these theory of mind changes is that they may contribute to the prerequisites that make children ready for formal instruction (Astington & Pelletier, 2005; Ziv & Frye, 2004). The conventional notion of school readiness emphasizes children's cognitive ability and their specific skills in literacy, numeracy, and social understanding. In addition, the current study suggests that 5-year-olds achieve a better understanding of the *process* that underlies most of the formal schooling experience, namely, teaching. Consequently, they should be in a better position to realize what schooling is about or, in other words, understand its basic intent and be able to take an active part in direct instruction.

References

Ashley, J., & Tomasello, M. (1998). Cooperative problem-solving and teaching in preschoolers. *Social Development*, 7, 143–163.

Astington, J. W. (1991). Intention in the child' theory of mind. In D. Frye & C. Moore (Eds.), *Children's theories of mind* (pp. 157–172). Hillsdale, NJ: Erlbaum.

Astington, J. W., & Lee, E. (1991). *What do children know about intentional causation?* Paper presented at biennial meeting of the Society for research in Child Development, Seattle, WA.

Astington, J. W., & Pelletier, J. (1996). The language of mind: Its role in teaching and learning. In D. R. Olson & N. Torrance (Eds.), *The handbook of education and human development: New models of learning, teaching, and schooling* (pp. 593–619). Cambridge, MA: Blackwell.

Astington, J. W., & Pelletier, J. (2005). Theory of mind, language, and learning in the early years: Developmental origins of school readiness. In B. D. Homer & C. Tamis-Lemonda (Eds.), *The development of social cognition and communication* (pp. 205–230). Mahwah, NJ: Lawrence Erlbaum.

Bartsch, K., Horvath, K., & Estes, D. (2003). Young children's talk about learning events. *Cognitive Development*, *18*, 177–193.

Beal, C. (1988). Children's knowledge about representations of intended meaning. In J. W. Astington, P. L.

Harris, & D. R. Olson (Eds.), *Developing theories of mind* (pp. 315–325). New York: Cambridge University Press.

Caro, T. M., & Hauser, M. D. (1992). Is there teaching in nonhuman animals? *Quarterly Review of Biology, 67*, 151–174.

Clement, F., Koenig, M., & Harris, P. (2004). The ontogenesis of trust. *Mind & Language, 19*, 360–379.

Davis-Unger, A. (2008). Development of teaching skills and relations to theory of mind in preschoolers. *Journal of Cognition and Development, 9*, 26–45.

Donaldson, M. (1978). *Children's minds*. Glasgow, Scotland: Fontana.

Feinfield, K. A., Lee, P. P., Flavell, E. R., Green, F. L., & Flavell, J. H. (1999). Young children' understanding of intention. *Cognitive Development, 14*, 463–486.

Flavell, J. H., & Miller, P. H. (1998). Social cognition. In D. Kuhn & R. S. Siegler (Eds.) & W. Damon (Series Ed.), *Handbook of child psychology: Vol. 2. Cognition, perception and language* (pp. 851–898). New York: Wiley.

Frye, D. (1999). Development of intention: The relation of executive function to theory of mind. In P. D. Zelazo, J. W. Astington, & D. R. Olson (Eds.), *Developing theories of intention: Social understanding and self-control* (pp. 119–132). Mahwah, NJ: Lawrence Erlbaum.

Frye, D., & Ziv, M. (2005). Teaching and learning as intentional activities. In B. D. Homer & C. Tamis-Lemonda (Eds.), *The development of social cognition and communication* (pp. 231–258). Mahwah, NJ: Lawrence Erlbaum.

Jeffrey, B. (2003). Countering learning 'instrumentalism' through creative mediation. *British Educational Research Journal, 29*, 489–503.

Karniol, R. (1978). Children's use of intention cues in evaluating behavior. *Psychological Bulletin, 85*, 76–85.

Koenig, M., & Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development, 76*, 1261–1277.

Kruger, A. C., & Tomasello, M. (1996). Cultural learning and learning culture. In D. R. Olson & N. Torrance (Eds.), *The handbook of education and human development* (pp. 369–387). Malden, MA: Blackwell Publishing.

Kuhn, D. (2000). Theory of mind, metacognition, and reasoning: A life-span perspective. In P. Mitchell & K. J. Riggs (Eds.), *Children's reasoning about the mind* (pp. 301–322). Hove, UK: Psychology Press.

Macmillan, C. J. B., & McClellan, J. E. (1968). Can and should means-ends reasoning be used in teaching? In C. J. B. Macmillan & T. W. Nelson (Eds.), *Concepts of teaching: Philosophical essays*. Chicago: Rand McNally.

Malle, B. F., Moses, L. J., & Baldwin, D. A. (2001). *Intentions and intentionality: Foundations of social cognition*. Cambridge, MA: MIT Press.

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? *American Psychologist, 59*, 14–19.

Miller, S. A. (2000). Children's understanding of preexisting differences in knowledge and belief. *Developmental Review, 20*, 227–282.

Montgomery, D. E. (1992). Young children's understanding of knowing: The development of a folk epistemology. *Developmental Review, 12*, 410–430.

Moses, L. J. (1993). Young children's understanding of belief constraints on intention. *Cognitive Development, 8*, 1–25.

Olson, D. R., & Bruner, J. (1996). Folk psychology and folk pedagogy. In D. R. Olson & N. Torrance (Eds.), *Handbook of education and human development* (pp. 9–27). Cambridge, MA: Blackwell.

Perner, J. (1988). Higher order beliefs and intentions in children's understanding of social interaction. In J. W. Astington, P. L. Harris, & D. R. Olson (Eds.), *Developing theories of mind* (pp. 271–294). New York: Cambridge University Press.

Perner, J., & Wimmer, H. (1985). "John thinks that Mary thinks that..." Attribution of second-order beliefs by 5- to 10-year-old children. *Journal of Experimental Child Psychology, 39*, 437–471.

Pillow, B. H. (1989). Early understanding of perception as a source of knowledge. *Journal of Experimental Child Psychology, 47*, 116–129.

Pillow, B. H. (1999). Children's understanding of inference and knowledge. *Journal of Genetic Psychology, 160*, 419–428.

Povinelli, D. J., & deBlois, S. (1992). Young children's (Homo sapiens) understanding of knowledge formation in themselves and others. *Journal of Comparative Psychology, 106*, 228–238.

Sabbagh, M. A., & Baldwin, D. A. (2001). Learning words from knowledgeable versus ignorant speakers: Links between preschooler's theory of mind and semantic development. *Child Development, 72*, 1054–1070.

Schult, C. A. (2002). Children's understanding of the distinction between intentions and desires. *Child Development, 73*, 1727–1747.

Shultz, T. R., & Wells, D. (1985). Judging the intentionality of action-outcomes. *Developmental Psychology, 21*, 83–89.

Sobel, D. M., Li, J., & Corriveau, K. (2007). "They danced around in my head and I learned them": Children's developing conceptions of learning. *Journal of Cognition and Development, 8*, 345–369.

Sodian, B., & Wimmer, H. (1987). Children's understanding of inference as a source of knowledge. *Child Development, 58*, 424–433.

Strauss, S., Ziv, M., & Stein, A. (2002). Teaching as a natural cognition and its relation to preschoolers' developing theory of mind. *Cognitive Development, 17*, 1473–1787.

Taylor, M. (1988). Conceptual perspective taking: Children's ability to distinguish what they know from what they see. *Child Development, 59*, 703–718, 786.

Taylor, M., Esbensen, B. M., & Bennett, R. T. (1994). Children's understanding of knowledge acquisition: The tendency for children to report that they have always known what they have just learned. *Child Development, 65*, 1581–1604.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, *72*, 655–684.

Wellman, H. M., & Woolley, J. (1990). From simple desires to ordinary beliefs: The early development of everyday psychology. *Cognition*, *35*, 245–275.

Wimmer, H., Hogrefe, J., & Sodian, B. (1988). A second stage in children's conception of mental life: Understanding informational accesses as origins of knowledge and belief. In J. W. Astington, P. L. Harris, & D. R. Olson (Eds.), *Developing theories of mind* (pp. 173–192). New York: Cambridge University Press.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*, 103–128.

Wood, D., Wood, H., Ainsworth, S., & O'Malley, C. (1995). On becoming a tutor: Toward and ontogenetic model. *Cognition and Instruction*, *13*, 565–581.

Zelazo, P. D., Astington, J. W., & Olson, D. R. (1999). *Developing theories of intention: Social understanding and self-control*. Mahwah, NJ: Lawrence Erlbaum.

Ziv, M., & Frye, D. (2004). Children's understanding of teaching: The role of knowledge and belief. *Cognitive Development*, *19*, 457–477.

Appendix

Protocols for Stories in Study 1

Successful teaching. Here are Itamar and Esti. Esti does not know how to make a bowl from clay. Itamar knows how to make a bowl from clay. Every day Itamar shows Esti how to make a bowl from clay, so that he learns how to do it. "Look, first you role the clay to a ball, then you stick your thumb into the ball, there, you have a bowl." Now Esti also knows how to make a bowl.

> Control Question 1: "At the beginning did Esti know how to make a bowl or didn't she know?"
> Control Question 2: "Does Esti know now how to make a bowl or doesn't she know how?"
> Test Question: "Did Itamar try to teach Esti how to make a bowl from clay or didn't he try to teach him?"

Successful imitation. Here are Uri and Dorit. Uri does not know how to tie his shoe laces. Dorit knows how to tie shoe laces. Every day Uri watches Dorit when she ties her shoe laces. (*Experimenter enacts Uri hiding behind Dorit.*) See, he sits here, behind the table, watches Dorit, and tries to do just what she does. Dorit does not see Uri. Dorit does not know that Uri is watching her. Look, now Uri knows how to tie his shoelaces.

> Control Question 1 (Perceptual Access Question): "Did Dorit know that Uri was watching her or didn't she know he was watching her?"
> Control Question 2: "At the beginning, did Uri know how to tie his shoelaces or didn't he know how?"
> Control Question 3: "Does Uri know now how to tie his shoelaces or doesn't he know how?"
> Test Question: "Did Dorit try to teach Uri how to tie his shoelaces or didn't he try to teach her?"

Failed teaching. Here are Gal and Ronit. Gal does not know how to play Animal Dominos. Ronit knows how to play Animal Dominos. Every day Ronit shows Gal how to play Animal Dominos, so that he learns how to play. "See, first you look for the same picture as here, and then you put the picture right next to this one. There, now you have two pictures that are the same right next to each other." Gal tried and tried to play, but he still does not know how to play. (*Experimenter enacts Gal putting pictures next to the wrong target picture.*)

> Control Question 1: "At the beginning did Gal know how to play Animal Dominos or didn't he know how to play?"
> Control Question 2: "Does Gal know now how to play Animal Dominos or doesn't he now know how?"
> Experimental Question: "Did Ronit try to teach Gal how to play Animal Dominos or didn't she try to teach him?"

Failed imitation. Here are Ofer and Galia. Ofer does not know how to jump on one leg. Galia knows how to jump on one leg. Every day Ofer watches Galia when she jumps on one leg, and he also tries to jump on one leg. See, he stands here, behind the door, watches Galia, and tries to do just what she does. Galia does not see Ofer. Galia does not know that Ofer is watching her. Ofer tries again and again to jump on one leg, but he still does not know how to do it. (*Experimenter enacts Ofer falling.*)

> Control Question 1 (Perceptual Access Question): "Did Galia know that Ofer was watching her or didn't she know he was watching her?"
> Control Question 2: "At the beginning did Ofer know how to jump on one foot or didn't he know how?"
> Control Question 3: "Does Ofer know now how to jump on one foot or doesn't he know how?"
> Test Question: "Did Galia try to teach Ofer how to jump on one foot or didn't she try to teach him?"

Unknown outcome of teaching. Here are Dan and Shani. Dan does not know how to write numbers. Shani knows how to write numbers. Every day Shani shows Dan how to write numbers, so that he learns how to write them. "See, this is how you write '1' and this is how you write '2'. Here are '1' and '2'."

> Control Question 1: "At the beginning did Dan know how to write numbers or didn't he know how to write them?"
> Test Question: "Did Shani try to teach Dan how to write numbers or didn't she try to teach him?"

Unknown outcome of imitation. Here are Dana and Amit. Dana does not know how to draw a house. Amit knows how to draw a house. Every day Dana watches Amit when he draws a house, and she also tries to draw a house. See, she sits here, behind the wall, watches Amit, and tries to do just what he does. Amit does not see Dana. Amit does not know that Dana is watching him.

> Control Question 1 (Perceptual Access Question): "Does Amit know that Dana is watching him or doesn't he know she is watching him?"
> Control Question 2: "At the beginning did Dana know how to draw a house or didn't he know how to draw a house?"
> Test Question: "Did Amit try to teach Dana how to draw a house or didn't he try to teach her?"

Standard false-belief stories. Here is Eyal and here is Tami. Eyal has a block. Eyal puts the block in the box and then he goes out to play. Now Eyal can't see and can't hear what Tami is doing. Tami takes the block out of the box and puts the block in the drawer instead.

> Control Question 1: "Where is the block now, is it in the box or is it in the drawer?"
> Control Question 2: "Where did Eyal put the block before going outside, did he put it in the box or did he put it in the drawer?"
> Test Question: "When he comes back, where will Eyal look for the block? Will he look in the box or will he look in the drawer?"

Protocols for New Stories in Study 2

Partial teaching. Here are Inbar and Oren. Oren wants to know how to write his name, but he doesn't know to write any letters. Inbar knows how to write. Look, here she's writing the names "Inbar" and "Oren." Every day Inbar shows Oren how to write his name, so that he knows how to write it. "Look, first you have to write the letter 'O' and then 'r', 'e', and then the letter 'n', like this." Oren tried to write his name but he can only write the two letters 'o' and 'r', like this.

> Control Question 1: "At the beginning did Oren know how to write his name or didn't he know how to write his name?"
> Control Question 2: "Does Oren know now how to write his name or doesn't he know how to write his name?"
> Control Question 3: "Does Oren know now to write some letters of his name or doesn't he know how to write some letters of his name?"
> Test Question: "Did Inbar try to teach Oren how to write his name or didn't she try to teach him?"

Partial imitation. Here are Roee and Yael. Yael wants to know how to draw a face of a man, but she doesn't know how to draw faces. Roee knows to draw a face. Every day Yael watches Roee drawing a face of a man. (*Experimenter enacts Yael hiding behind Roee.)* See, she sits here, behind the table, watches Yael, and tries to do just what he does. Roee does not see Yael. Roee does not know that Yael is watching him. Now Yael tries to draw a face of a man but she can only draw a circle with hair, like this.

> Control Question 1 (Perceptual Access Question): "Did Roee know that Yael was watching him or didn't he know she was watching him?"
> Control Question 2: "At the beginning did Yael know how to draw a face of a man or didn't she know how to draw a face of a man?"
> Control Question 3: "Does Yael know now how to draw a face of a man or doesn't she know how to draw a face of a man?"
> Control Question 4: "Does Yael know now to draw part of the face or doesn't she know how to draw part of a face of a man?"
> Test Question: "Did Roee try to teach Yael how to draw a face of a man or didn't he try to teach her?"

Hidden intention game. Here are Dan and Naomi. Dan wants to read numbers, but he doesn't know how to read numbers at all. Naomi knows how to read numbers. Look, here she is reading numbers 1, 2, 3, all the way to 10. Every day Naomi shows Dan how to read numbers, so that he knows how to read numbers too. "Look, this is number 1, this is 2, and this is number 3..." Today Naomi brings a game to class. Naomi says to Dan, "We are going to play the number game now. In this game every child gets a card and has to say what number appears on the card. The winner is the one who reads all the numbers.'' Now Naomi and Dan are playing the number game.

> Test Question: ''When the Naomi plays the number game with Dan, what does she really want? Does she want to play with the game with Dan or does she want to teach him numbers?''

Hidden intention question. Here are Tamar and Adam. Tamar wants to know names of shapes, but she doesn't know any name of any shape. Adam knows the names of all shapes. Look, here he's pointing at shapes and says, ''this is a square and this is a circle.'' Every day Adam shows Tamar shapes and he tells her what is the name of each shape, so that she knows its names too. ''Look, this is a triangle, and this is a circle.'' Now Adam shows Tamar a card with a shape drawn on it and asks her, ''What is drawn here?''

> Test Question: ''Why does Adam asks Tamar this question, because he wants to know what shape is drawn on the card or because he wants to see whether Tamar knows what shape is drawn on the card?''

Intention unfulfilled/desire satisfied (adapted from Schult, 2002). Here is Batya. Batya really likes playing with dolls. She would like a new doll to play with. There's a doll at the toy store that Batya wants. Batya has enough money to buy the doll. She decides she's going to the toy store to buy the doll. That day, before she went to the store, her mother gave her the doll. Now Batya has the doll.

> Control Question 1: ''At the beginning, did Batya have the doll or did she not have the doll?''
> Control Question 2: ''Does Batya have the doll now or does she not have the doll now?''
> Test Question: ''Did Batya try to go to the store and buy the doll or did she not try to go and buy the doll?''

Intention fulfilled/desire unsatisfied (adapted from Schult, 2002). Here is Gil. It's a rainy day and Gil has to walk home from school. He wants his shoes to stay dry, because he doesn't like having wet shoes. He decides he's going to walk around all the puddles. Gil walks very carefully and doesn't step into a single puddle. When he gets home, his shoes are still dry. Just then, his little brother spills water on is shoes.

> Control Question 1: ''At the beginning were Gil's shoes wet or dry?''
> Control Question 2: ''Now, are Gil's shoes wet or dry?''
> Test Question: ''Did Gil try to walk around the puddles to keep his shoes dry or did he not try to walk around the puddles to keep his shoes dry?''

Child Development, September/October 2008, Volume 79, Number 5, Pages 1257 – 1269

Changes in the Relation of Self-Efficacy Beliefs and Behaviors Across Development

Pamela E. Davis-Kean, L. Rowell Huesmann, and Justin Jager
University of Michigan

W. Andrew Collins
University of Minnesota

John E. Bates
Indiana University

Jennifer E. Lansford
Duke University

Many social science theories that examine the connection between beliefs and behaviors assume that belief constructs will predict behaviors similarly across development. Converging research implies that this assumption may not be tenable across all ages or all belief constructs. Thus, to test this implication, the relation between behavior and beliefs about the self was examined in 2 independent data sets with 2 different constructs: aggression and achievement. The respondents were 6 – 18 years of age and predominately Caucasian. Results using quasi-simplex structural equation models suggest that self-beliefs become more strongly related to behavior as children grow older independent of the reliability of the measures used. Possible limitations in the use of self-report methodology with young children are discussed.

Children entering school are confronted with two important tasks: to achieve and to get along cooperatively with others. Because of the importance of these two tasks for positive development, many researchers have focused their research on trying to understand what might be the major influences on these outcomes. An important tenet of this work is that a link exists between beliefs about ability to perform a behavior (self-efficacy) and actually performing the behavior (e.g., Bandura, 1986; Crick & Dodge, 1994; Huesmann, 1988; Kohlberg, 1969; Wigfield & Eccles, 1992). The theoretical models based on these tenets are used to explain children's behaviors by the various belief domains that may be specific to the behavior. For example, when trying to understand achievement outcomes, children's beliefs about their ability to perform are often assessed and used to predict behavior (Bandura, 1986; Wigfield & Eccles, 1992). A similar method is used in research in which children's beliefs about appropriate behavior with peers are assessed in connection with their aggressive behaviors (Crick & Dodge, 1994; Huesmann & Guerra, 1997).

The models guiding this research imply that children's beliefs and expectations about their behavior will dictate or predict their behaviors and that this prediction is similar across development. Younger children, for example, would show similar relations between their beliefs about achievement and actual achievement as would older children. An assumption from these theories, then, is that a primary intervention for changing children's behaviors in achievement and peer relations is to change their beliefs about important behaviors and that the age of the child would not be a factor in this intervention. Even though these models linking beliefs and behaviors have been well examined across the years, the general focus has been on cross-sectional relations or short-term longitudinal studies with very limited age ranges (e.g., middle childhood or adolescence). This limitation primarily reflects the lack of longitudinal data that spanned from early schooling (e.g., first/second grade) to later schooling (high school). Thus, little research has examined the specific relation between self-beliefs and their associated behaviors across time.

This research was supported by the Center for the Analysis of Pathways from Childhood to Adulthood, which is funded by National Science Foundation Grant 0322356. We thank the following people for contributions of their data sets to this study: Kenneth A. Dodge and Gregory S. Pettit (Child Development Project), and Jacque Eccles, Alan Wigfield, Rene Harold, and Phyllis Blumenfeld (Childhood and Beyond Study). We are also very grateful for the statistical assistance of Chris Mutch, Jeremy Welland, Holly Sexton, and Jackson Goodnight.

Correspondence concerning this article should be addressed to Pamela E. Davis-Kean, 426 Thompson Street, University of Michigan, Ann Arbor, MI 48106-1248. Electronic mail may be sent to pdakean@umich.edu.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7905-0004

A primary goal of this article is to examine the relation between self-efficacy beliefs and behaviors across time to determine the stability of this relation. A question of particular interest is whether this relation remains stable across time or changes as children age. A related question is whether, if the belief–behavior relation changes, the change is related to age. A second goal of this article is to examine these relations for two important developmental outcomes: achievement and negative peer relations (aggression). This type of research is important for understanding whether self-referent beliefs about behaviors are similarly predictive across time. If self-referent beliefs do not relate to behaviors similarly across time, then it is important to understand how these relations may change so that interventions may be appropriately targeted for the age of the child.

The Development of Beliefs and Behaviors

Considerable research has suggested that younger children have difficulties with certain tasks and concepts (e.g., future expectations, predictions of others' behaviors, conservation tasks) but that these difficulties are less apparent at older ages (Alvarez, Ruble, & Bolger, 2001; Case, 1991; Harter, 1996; Parsons & Ruble, 1977; Ruble, Parsons, & Ross, 1976). Piaget (1963/2001; Siegler & Alibali, 2004) attributed these changes to a shift from the preoperational stage of reasoning to a concrete operational stage, which involves more abstract thinking and use of symbolic concepts beginning at about age 7. Although previous cognitive development research focused on the idea that children develop through discrete, qualitative stages, researchers currently emphasize that children steadily become better and more efficient information processors as the brain develops a greater capacity for more complicated procedures (e.g., parallel processing; Siegler & Alibali, 2004). Evidence of mental capabilities that children did not possess at earlier ages continues well into adolescence (Keating, 2004).

This revised view does not mean, however, that development is linear. For example, some dramatic changes appear to occur between the ages of 5 and 8, making this an important focus for research (Siegler & Alibali, 2004). Indeed, 8-year-olds have become the most commonly used age group in studies of children's ability to reason about themselves and others (Alvarez et al., 2001; Harter, 1996). Researchers have referred to this change as the "5–7 year shift" (Sameroff & Haith, 1986; White, 1970). White (1996), for example, referred to the time period after the age of 7 as entering the "age of reason." Historically, ages 5–7 have been regarded as an important turning point distinguishing childhood from early adulthood because of the changes in the mental ability of the child. However, there is also research to suggest that adolescence may be a key time of change as children's cognitive development continues to mature and shifts to reasoned action and reflection (Keating, 2004). It is important when examining children's beliefs to know whether or not significant shifts occur in children's ability to understand and then report on their beliefs across time in order to understand when these beliefs might influence behavior.

Development of Belief–Behavior Relations

Though few direct tests exist of the developmental change in belief–behavior relations, differences in belief–behavior connections before and after 8 years of age have been reported frequently. Findings from developmental research suggest that younger children's beliefs and behaviors may be only negligibly related before the age of 7 (Parsons & Ruble, 1977; Spencer & Bornholt, 2003) and that beliefs are unreliable predictors of future behaviors until after the age of 8 (Huesmann & Guerra, 1997). This phenomenon has been documented in research on such diverse domains as self-esteem (Davis-Kean & Sandler, 2001), expectancies regarding success and failure (Parsons & Ruble, 1977), beliefs about the acceptability of aggression (Huesmann & Guerra, 1997), and social information processing in relation to externalizing behavior (Lansford et al., 2006). These arrays of findings give some indication that the relations between self-efficacy beliefs and behaviors may vary across development.

Further support comes from research on expectancies for success and failure. Parsons and Ruble (1977), for example, showed that older children incorporated information regarding their outcomes quickly and changed their predictions about future successes or failures accordingly, whereas children younger than age 8 typically did not. This is consistent with developmental work by Harter (1999) suggesting that possible cognitive limitations to young children's ability to answer these types of questions (self-reflective) at earlier ages.

These findings, though suggestive, do not constitute direct evidence on the question of whether stability or change occurs in the relation between self-efficacy beliefs and behaviors across development. In order to test for stability, one would need to examine the synchronous correlations between self-efficacy behavioral beliefs and their associated behaviors and see if they differ across time. If they do not significantly differ across time, then these relations are stable. Conversely, if the synchronous correlations do significantly differ across time, then these relations are changing and it will be important to specify at what age these relations may be changing.

The current study is intended to describe and test whether or not the relations between self-efficacy beliefs and behaviors stay stable or change across time. We examine these relations across time by using information from multiple data sets and multiple belief domains. If the relations change, are the relations weaker in childhood than in adolescence, as previous research appears to suggest or do the associations differ from one age period to another across early childhood, middle childhood, and adolescence? An important consideration in examining these changes is whether reliability of the measures is a factor in the change, such that younger children are less reliable in their answers, thus attenuating the correlations between beliefs and behaviors (Davis-Kean & Sandler, 2001; Trzesniewski, Donnellan, & Robins, 2003).

We use two different longitudinal data sets with observations collected both in childhood and in adolescence. In these data sets, we have information on two different behaviors (i.e., academic achievement and aggressiveness) and self-efficacy beliefs regarding these behaviors (e.g., self-concept of ability and self-efficacy of aggression). Thus, a unique feature of this study is that we will test the generalizability of change in associations between self-efficacy beliefs and behaviors both through analyzing two independent data sets and by using two different important childhood outcomes.

Method

The data sets for the study come from a collaboration of longitudinal researchers who have made their data sets available to answer and replicate developmental questions. The data sets selected for the analyses met two criteria. First, both had been guided by theories (e.g., information processing theories, expectancy-value theory) positing similar self-efficacy belief–behavior relations across development. Thus, the researchers had collected data that measured beliefs and behaviors on multiple occasions across childhood and adolescence. Second, the self-efficacy beliefs and behaviors were assessed in a similar manner across time. The two data sets that met these criteria and their relevant methodology are summarized below.

Data Sets and Measures

Childhood and Beyond (CAB). Three cohorts of children from 28 schools located in four southeastern Michigan school districts provided data for the current analysis (Eccles, Wigfield, Harold, & Blumenfeld, 1993). Academic belief (math self-concept of ability) and behavior (math grades) data were collected starting in the fall and spring of the 1988–1989 school year, respectively, from children in Grade 1 ($n = 317$), Grade 2 ($n = 330$), and Grade 4 ($n = 423$) each year for 3 years. In a follow-up study starting in 1994, data were collected from these cohorts each year for 3 years. Thus, the Grade 1 cohort has data from six periods (Grades 1, 2, 3, 7, 8, and 9), the Grade 2 cohort has data from six periods (Grades 2, 3, 4, 8, 9, and 10), and the Grade 4 cohort has data from six periods (Grades 4, 5, 6, 10, 11, and 12). Ninety-three percent of participants in the sample are Caucasian, 50% are female, and all generally were of relatively high socioeconomic status (SES).

Self-concept of math ability (SCMA) and math grades served as the primary belief and behavior variables of interest, respectively. SCMA is a composite variable that includes five items: "How good at math are you?" "If you had to list all the students from best to worst in math where are you?" "Compared to other subjects how good are you at math?" "How well do you expect to do in math this year?" and "How good would you be at learning something new in math?" Responses were measured on a Likert-type scale ranging from 1 to 7 (e.g., 1 = *one of the worst*, 7 = *one of the best*) and collected in the spring of the child's academic year. SCMA Cronbach's alphas were .76, .79, .85, .90, .92, and .93 for Waves 1 through 6, respectively.

Math grades were obtained from school records. In the first three waves, math grades correspond to the child's math performance during the spring only; in the latter three time periods, the child's final math grade for the academic year was used. Grades were coded on a scale of 1–16 where 1 represented an F− and 16 an A+. The grades ranged between 1 and 16 for the Grade 1 cohort, between 3 and 16 for the Grade 2 cohort, and between 1 and 16 for the Grade 4 cohort.

Child Development Project (CDP). The families in the CDP (Dodge, Bates, & Pettit, 1990) were recruited when the children entered kindergarten in 1987 or 1988 at three sites: Knoxville and Nashville, TN, and Bloomington, IN. The combined sample consisted of 585 families at the first assessment. Males comprised 52% of the sample. Eighty-one percent of the sample were Caucasian, 17% were African American, and 2% were from other ethnic groups. Follow-up assessments used in the present study were conducted in Grades 1–3, 8, and 11, giving a total of six waves of data. Compared to the original sample, the families who provided data in Grade 11 (78% of the original sample) were of slightly higher SES, but participants and nonparticipants did not differ by race, single-parent status, or mothers' reports of children's externalizing behaviors in kindergarten.

In Grades kindergarten through 3, children were presented with a series of four cartoon pictures and brief verbal descriptions of the cartoon events. In each cartoon, children were told to imagine being the protagonist. After hearing the description of each event, children were presented with alternative strategies (competent, aggressive, and inept) for dealing with the situation. Efficacy for aggression was assessed by asking children to rate how easy it would be for them to act aggressively in each situation (1 = *HARD*, 2 = *hard*, 3 = *easy*, 4 = *EASY*). Efficacy for aggression was scored as the average of this item following the aggressive response across the four vignettes. Cronbach's alphas for this measure in Grades 1–3 were .64, .59, and .55, respectively. In Grades 8 and 11, participants were shown six video vignettes that started with a social interaction and culminated in an ambiguous provocation by peers or adults directed toward a protagonist adolescent. Participants were asked to imagine being the protagonist. Three vignettes depicted provocations that were relevant to both boys and girls and were presented to all participants; the remaining three vignettes were sex specific. Each provocation segment was followed by two segments that presented an aggressive and a nonaggressive response to the provocation, respectively. After each of the responses presented on the video, efficacy for aggression was assessed by asking adolescents how easy it would be for them to act aggressively in each situation (1 = *very hard*, 5 = *very easy*). A composite variable was created in Grade 8 and Grade 11 by averaging this item from the aggressive responses across the six vignettes. Cronbach's alphas for these measures were .82 and .86, respectively.

Mothers completed the Child Behavior Checklist (CBCL; Achenbach, 1991) annually when their children were in kindergarten through grade 11. Mothers reported whether each item was 0 = *not true*, 1 = *somewhat or sometimes true*, or 2 = *very or often true* of the child. The 20 items in the aggression subscale were summed to create a separate scale for each year. Cronbach's alphas were high and stable (.87, .87, .87, .89, .89, .90 for Waves 1 through 6, respectively).

Attrition and Missing Data Plan

The problem of missing data is inherent in the use of longitudinal data sets. Each of the data sets used in this study had various amounts of missing data—ranging from a low amount of missing (CDP ~ 5%) to a rather high amount of missing (CAB ~ 40%). In order to minimize bias due to attrition, the CAB and CDP missing data were multiply imputed using the NORM program (Schafer, 1999). Compared to other missing data procedures, multiple imputation (MI) is superior when it comes to minimizing bias due to attrition (Graham, Hofer, & Piccinin, 1994; Little & Rubin, 1987; Schafer & Olsen, 1998). In the same regard, MI is also far superior to simple casewise deletion (Little & Rubin, 1987; Schafer & Olsen, 1998). Moreover, not only are the parameter estimates more trustworthy when using MI (due to its accounting for attrition more effectively) but also standard errors associated with the parameter estimates are not artificially decreased, and as such inferences are trustworthy as well (Graham et al., 1994; Schafer & Olsen, 1998).

One assumption of the MI procedure is that all variables included in the procedure are approximately normally distributed (Schafer & Olsen, 1998). Any data with violations of normality (e.g., aggressive behavior variables) were transformed prior to imputation. A second assumption of the MI procedure is that the data imputed are missing at random. When utilizing MI, the first step is to empirically identify as many variables as possible that are related to the missingness of the variables to be imputed. Once identified, these variables are included in the imputation process as "auxiliary variables," and their inclusion helps to ensure that the imputed data resulting from the MI procedure are unbiased. However, once the data are imputed, the auxiliary variables are dropped, and the imputed data are then analyzed separately from the auxiliary variables. Because the auxiliary variables are not included in the actual analyses, they are not equivalent to control variables.

Separately for CAB and CDP, the number of missing time points for each variable was used as a grouping variable, and a series of one-way analyses of variance were conducted to determine what auxiliary variables were significantly related to missingness. Several auxiliary variables were identified and incorporated into the MI model for both data sets, including SES indicators (i.e., income, education, and occupation), IQ measures (e.g., Slossan, Wechsler Intelligence Scale for Children–Revised), sex, race/ethnicity, and demographic location variables such as neighborhoods or school districts.

Due to the cohort-sequential design of CAB, there are many instances of planned missingness (e.g., a respondent is missing data due to the design of the study as opposed to nonresponse). In instances of planned missingness, data were not imputed. For example, due to the design of the study, no respondent in the first cohort of CAB provided data at Grades 4, 5, 6, 10, 11, and 12. Thus, individuals in the first cohort of CAB did not have data imputed at these grades, and instead only had data imputed for the grades where data were actually collected (Grades 1, 2, 3, 7, 8, and 9). In a similar fashion, for the remaining two CAB cohorts, data were imputed only for the grades where data were actually collected. For each respondent, data were imputed at the individual item level (as opposed to the scale level). This was done for three reasons. First, when using MI and imputing scale data, it is standard practice to carry out those imputations at the item level (Schafer & Olsen, 1998). Second, when possible, we model the multiitem scales as latent constructs in our structural equation models, and therefore, our analyses demand item-level data. Finally, as is standard practice when utilizing MI (Graham et al., 1994; Schafer & Olsen, 1998), in instances where respondents had complete missing data for a given measure at a particular grade (whether based on a single-item or a multiitem scale), and that complete missingness was not due to planned missingness, data were imputed.

Among the 1,070 respondents included in the CAB analyses, the mean percentage missing across the 36 measures for which missing values were imputed (five observations for the math self-concept of ability scale at six time points and one observation of grades at six time points) was 40.1%. Based on this fraction, eight data sets were imputed to reach an efficiency level above .95 (Rubin, 1987). Among the 585 respondents included in the CDP analyses, the mean percentage missing across the 148 measures for which missing values were imputed (20 observations for the CBCL at six time points, 4 observations for the efficacy for aggression scale at the four initial time points, and 6 observations for the efficacy for aggression scale at the last two time points) was 5.3%. Based on this fraction, only three data sets needed to be imputed in order to reach an efficiency level above .95 (Rubin, 1987); however, because five imputed data sets is the standard minimum for number of imputed data sets (Schafer & Olsen, 1998), five data sets were imputed to reach an efficiency level well above .95.

Analysis Strategy

Imputed means and estimates of their standard errors were calculated using PROC MIANALYZE within SAS (SAS Institute, 1999), which was developed to estimate standard errors of multiply imputed data. All analyses were conducted within Mplus (Muthen & Muthen, 1998–2006), using the imputation option developed for multiply imputed data.

In order to assess if the association between SCMA and math grades changes over time, we used the quasi-simplex measurement model presented in Figure 1. In order to asses if the relation between efficacy for aggression and aggressive behavior changes over time, we used the quasi-simplex measurement model presented in Figure 2. First, the models in Figures 1 and 2 are measurement models (as opposed to causal path models) because there are no causal paths in either model and, instead, only correlations are included. The inclusion of causal paths (say, autoregressive paths) would answer a different question—namely, "how does the relationship between changes in beliefs and changes in behaviors between time *t* and time *t*−*1* itself change over time?"

Second, the models in Figures 1 and 2 are both quasi-simplex models because in both models the behavior measure (i.e., math grade for the model in Figure 1 and aggressive behavior for the model in Figure 2) is based on a single indicator. Within a quasi-simplex model, in order to adjust for the reliability of each measure at each time point, errors are fixed based on the following equation: $e = \sigma^2[1 - \alpha]$ (Joreskog, 1970). For the model in Figure 1, the measure for behavior (i.e., final math grade) is based on a single indicator; as such the interitem reliability between fall math grade and final math grade was used as a proxy measure for interitem reliability (allowing for the fact that grades generally represent a true score with some error). For each of the six points in time, a given point in time's interitem reliability between fall math grade and final math grade was applied. Across the six time periods, the alphas were .85, .85, .88, .95, .96, and .96, respectively. For the model in Figure 2, the measure for behavior (i.e., aggressive behavior) is based on the aggregate of 20 CBCL indicators for which the interitem reliability at each time point was applied. Across the six periods, the alphas were .87, .87, .87, .89, .89, and .90, respectively. For both of the latent behavior constructs (math grades and aggressive behavior), it is not possible to correlate the error variances of matching items across time because they are based on a single observed indicator (Marsh, 1993). However, for both the latent belief constructs (SCMA and efficacy for aggression), which were based on multiple observed indicators, the error variances of matching items were correlated across time.

In order to empirically assess whether or not the relation between behaviors and beliefs changes over time, we compared models where the synchronous covariances between beliefs and behaviors are constrained to be equal across time to models where the synchronous covariances are free to vary across time.

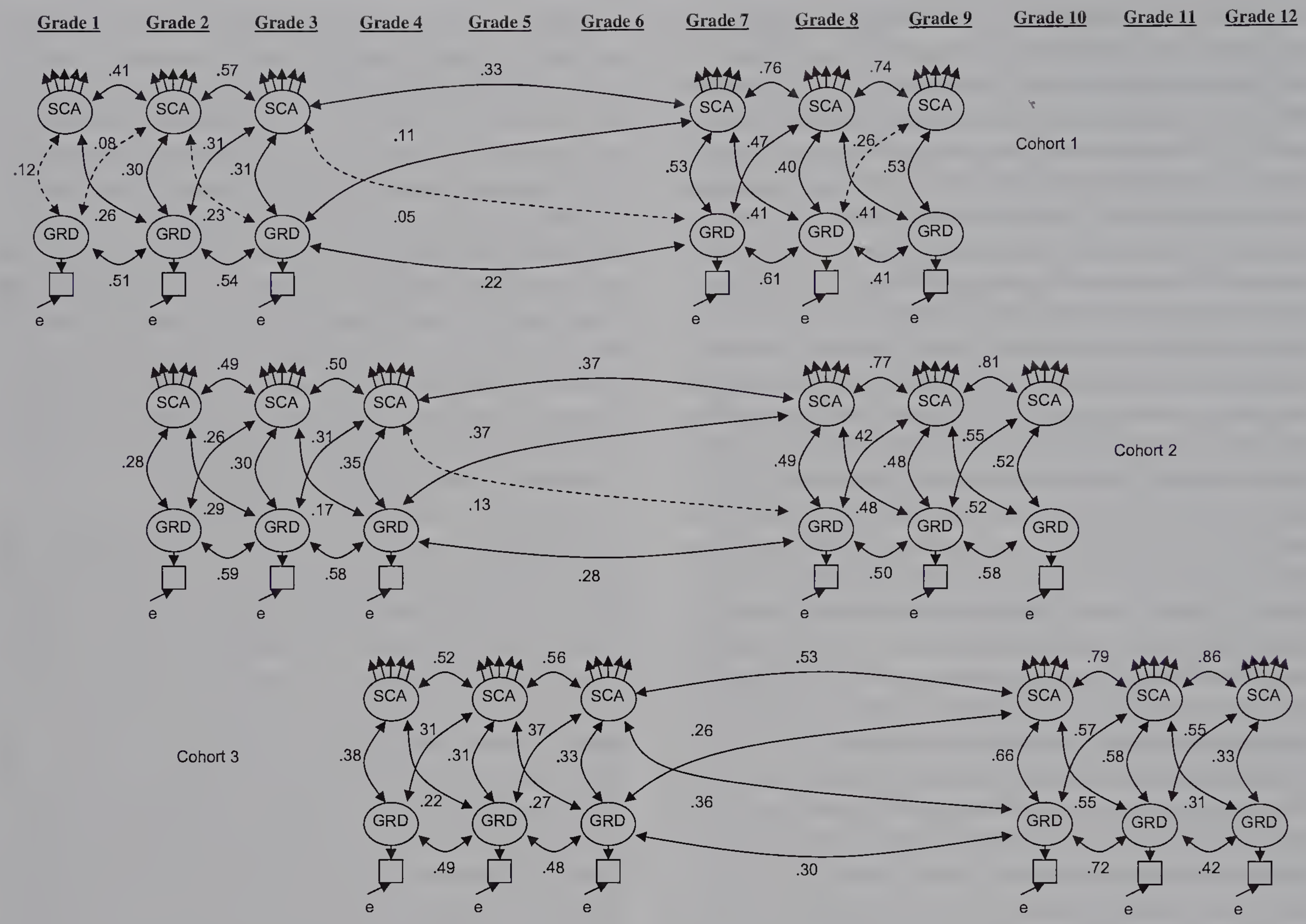

Grade 1
Grade 2
Grade 3
Grade 4
Grade 5
Grade 6
Grade 7
Grade 8
Grade 9
Grade 10
Grade 11
Grade 12
Cohort 1
SCA
GRD
e
.41
.57
.33
.76
.74
.08
.31
.11
.47
.26
.12
.30
.31
.53
.40
.53
.26
.23
.05
.41
.41
.51
.54
.22
.61
.41
Cohort 2
.49
.50
.37
.77
.81
.26
.31
.37
42
.55
.28
.30
.35
.49
.48
.52
.29
.17
.13
.48
.52
.59
.58
.28
.50
.58
Cohort 3
.52
.56
.53
.79
.86
31
37
.26
.57
.55
.38
.31
.33
.66
.58
.33
.22
.27
.36
.55
.31
.49
.48
.30
.72
.42

Figure 1. Estimates for quasi-simplex measurement model assessing relationship between math self-concept and math grades over time, by cohort, Childhood and Beyond Study.
Note. Important correlations are bolded for emphasis. Solid lines indicate significance at .05 level. SCA self-concept of ability; GRD short for math grade. For SCA, exogenous arrows represent indicators listed in Table 2. Fit indices: $\chi^2(1,425) = 3,395.58, p < .001$; comparative fit index = .92; root mean square error of approximation = .06.

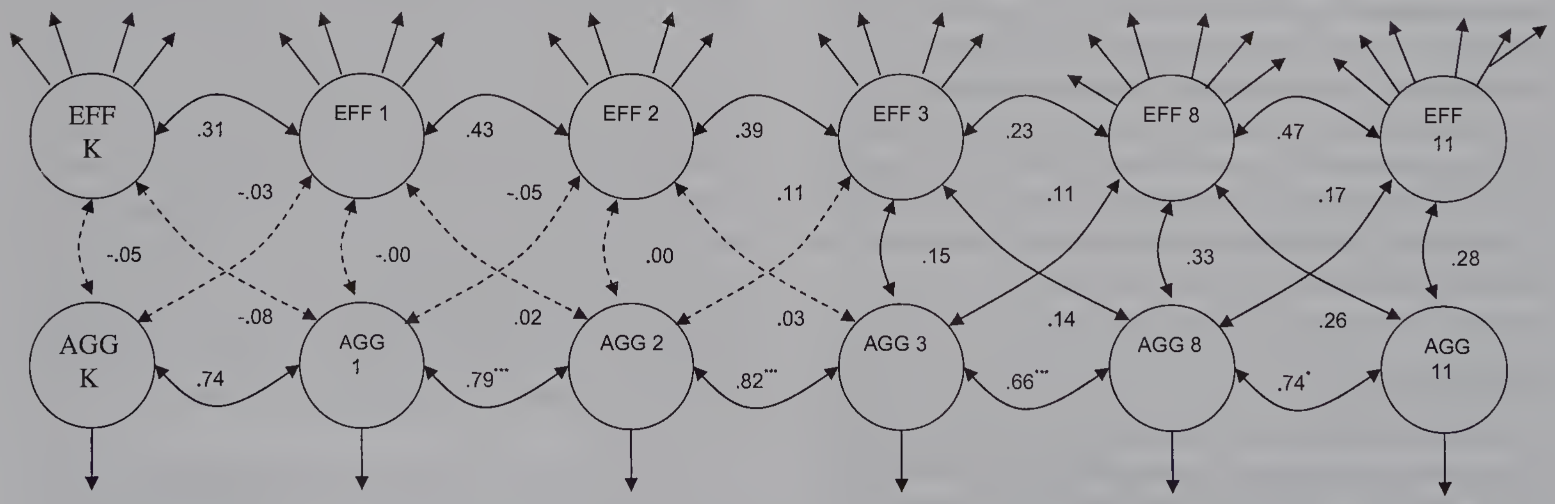

EFF K
EFF 1
EFF 2
EFF 3
EFF 8
EFF 11
.31
.43
.39
.23
.47
-.03
-.05
.11
.11
.17
-.05
-.00
.00
.15
.33
.28
-.08
.02
.03
.14
.26
AGG K
AGG 1
AGG 2
AGG 3
AGG 8
AGG 11
.74
.79***
.82***
.66***
.74*

Figure 2. Estimates for quasi-simplex measurement model assessing relationship between aggression self-efficacy and aggression over time, Child Development Project.
Note. Important correlations are bolded for emphasis. Solid lines indicate significance at .05 level. EFF short for aggression self-efficacy; AGG short for aggression. For EFF, exogenous arrows represent indicators listed in Table 2. Fit indices: $\chi^2(922) = 1,682.52, p < .001$; comparative fit index = .90; root mean square error of approximation = .05.

According to convention, model comparisons were based on changes in chi-square relative to changes in degrees of freedom (Kline, 1998). The chi-square value used in the model comparisons is the chi-square value provided in the Mplus output, which is the average chi-square value (Muthen, 2005) across the *m* data sets, where m is the number of imputed data sets.

Results

Basic Descriptive Statistics

The imputed means and standard deviations for all measures for CAB and CDP are presented in Table 1. The means and standard deviations are presented separately for each cohort. The factor loadings for the latent belief constructs, math self-concept of ability (CAB), and efficacy for aggression (CDP), and the latent behavior constructs, math grades, and aggressive behavior, were computed from the measurement models in Figures 1 and 2 and are listed in Table 2. Also listed in Table 2 are the residual variances for each observed variable included in the measurement models in Figures 1 and 2. Because there are eight imputed data sets for CAB and five imputed data sets for CDP, there are 13 covariance matrices in all. As such the covariance matrices are not presented here but are available from the authors upon request.

Changes in Synchronous Correlations Between Math Self-Concept and Math Grades

Before examining whether or not the association between math self-concept and math grades shown in Figure 1 changes over time, we tested for cohort and sex differences in the relations shown in the model. First, we tested for cohort differences. Specifically, a fully unconstrained model (Model 1, Table 3) where all covariances were free to vary across the three cohorts was compared to both a model where just the synchronous covariances were constrained to be equal across the three cohorts (Model 2, Table 3) and a model where all covariances were constrained to be equal across the three cohorts (Model 3, Table 3). The χ^2_{diff} test showed that these models differed significantly from each other, and therefore, it was not tenable that these associations were same across the cohorts, Model 2, $\chi^2_{\text{diff}}(12) = 60.37, p < .001$, and Model 3, $\chi^2_{\text{diff}}(52) = 197.19, p < .001$. Next, we tested for sex differences. Again, we compared a fully unconstrained model (Model 4, Table 3) where all covariances were free to vary across sex to both a model where just the synchronous covariances were constrained to be equal across sex (Model 5, Table 3) and a model where all covariances were constrained to be equal across sex (Model 6, Table 3). The fit of Model 4 (the fully unconstrained model) did not fit the data any better than did Model 5, $\chi^2_{\text{diff}}(6) = 7.96, p = .24$, or

Table 1
Overall Means for CAB and CDP, by Cohort

		Point of assessment					
	N	Time 1	Time 2	Time 3	Time 4	Time 5	Time 6
CAB							
Cohort 1	317						
Math SCA		5.62 (1.03)	5.53 (0.98)	5.54 (1.00)	4.94 (1.25)	4.69 (1.28)	4.60 (1.37)
Math grade		9.96 (2.01)	10.52 (1.91)	10.79 (1.98)	11.95 (2.78)	11.49 (2.67)	10.06 (3.42)
Cohort 2	330						
Math SCA		5.49 (0.67)	5.45 (0.69)	5.32 (0.73)	5.00 (0.82)	4.76 (0.89)	4.73 (0.94)
Math grade		10.65 (1.36)	10.99 (1.47)	11.87 (1.13)	11.97 (2.00)	11.00 (2.36)	11.13 (2.23)
Cohort 3	423						
Math SCA		5.35 (0.76)	5.35 (0.78)	5.25 (0.82)	4.82 (0.93)	4.66 (1.01)	4.54 (1.07)
Math grade		11.11 (1.54)	12.04 (1.67)	11.99 (1.28)	10.43 (2.26)	9.58 (2.67)	9.92 (2.53)
CDP							
Cohort 1	308						
Aggression efficacy		0.48 (0.32)	0.43 (0.30)	0.42 (0.33)	0.40 (0.33)	0.39 (0.32)	0.32 (0.35)
CBCL		2.64 (0.86)	2.45 (0.90)	2.38 (0.97)	2.14 (1.07)	2.08 (1.32)	1.93 (1.11)
Cohort 2	277						
Aggression efficacy		0.46 (0.30)	0.41 (0.30)	0.38 (0.30)	0.37 (0.32)	0.35 (0.32)	0.33 (0.37)
CBCL		2.45 (0.85)	2.31 (0.95)	2.13 (0.98)	2.07 (1.33)	2.19 (1.35)	2.05 (1.21)

Note. Standard deviations are in parentheses. CAB = Childhood and Beyond Study; CDP = Child Development Project; CBCL = Child Behavior Checklist; SCA = self-concept of ability.

Table 2

Factor Loadings for Math Self-Concept of Ability (CAB), Math Grade (CAB), Efficacy of Aggression (CDP), and Aggression (CDP)

	Point of assessment					
	Time 1	Time 2	Time 3	Time 4	Time 5	Time 6
Math self-concept of ability (CAB)						
How good at math are you?	.776 (.41)	.847 (.29)	.858 (.27)	.903 (.21)	.921 (.18)	.910 (.19)
If you had to list all the students from best to worst in math where are you?	.720 (.49)	.782 (.40)	.829 (.32)	.749 (.48)	.779 (.44)	.779 (.42)
Compared to other subjects how good are you at math?	.490 (.76)	.620 (.62)	.664 (.57)	.839 (.33)	.862 (.30)	.841 (.32)
How well do you expect to do in math this year?	.670 (.56)	.593 (.66)	.710 (.51)	.762 (.46)	.712 (.54)	.829 (.33)
How good would you be at learning something new in math?	.489 (.77)	.430 (.82)	.574 (.68)	.726 (.51)	.796 (.41)	.820 (.35)
Math grade (CAB)	.926 (.14)	.924 (.15)	.942 (.11)	.977 (.05)	.981 (.04)	.982 (.04)
Efficacy for aggression (CDP)						
Pushing kids out of line is hard for you?	.462 (79)	.514 (.74)	.552 (.70)	.664 (.56)		
Telling a kid to stop changing the channel or you will hit him/her is hard for you?	.723 (.48)	.666 (.56)	.603 (.64)	.750 (.44)		
Telling a kid that he/she had better let you play is hard for you?	.622 (.61)	.661 (.56)	.742 (.45)	.766 (.41)		
Telling a kid that you will hit him/her if he/she does not let you see his/her photos is hard for you?	.608 (.63)	.682 (.53)	.697 (.51)	.708 (.50)		
How easy would it be for you to act like this? (Vignette 1)					.384 (.85)	.653 (.57)
How easy would it be for you to act like this? (Vignette 2)					.602 (.63)	.744 (.44)
How easy would it be for you to act like this? (Vignette 3)					.550 (.70)	.721 (.48)
How easy would it be for you to act like this? (Vignette 4)					.764 (.42)	.707 (.50)
How easy would it be for you to act like this? (Vignette 5)					.766 (.41)	.595 (.65)
How easy would it be for you to act like this? (Vignette 6)					.674 (.54)	.747 (.44)
Child Behavior Checklist (CDP)	.933 (.13)	.933 (.13)	.932 (.13)	.944 (.11)	.945 (.11)	.946 (.11)

Note. Residual variances in parentheses. CAB = Childhood and Beyond Study; CDP = Child Development Project.

Model 6, $\chi^2_{\text{diff}}(26) = 35.18$, $p = .11$, and the χ^2_{diff} test did not significantly differ from the unconstrained model. Thus, we can conclude that the associations between math self-efficacy and math grades did not differ across sex.

Given these findings, we proceeded to estimate the simplex model of relations between math self-efficacy and math grades across grades separately for each cohort with the parameters constrained to be equal across males and females. The resulting parameter estimates are shown in Figure 1 (note that the standardized estimates or correlations are presented opposed to the unstandardized estimates or covariances). The estimates for the synchronous correlations between math self-concept and math grades are of key interest. In order to test if the synchronous correlations vary over time, we conducted a series of model comparisons. First, we compared a model where all synchronous covariances were free to vary across time (Model 7, Table 3) to a model where all synchronous covariances were constrained to be equal across time (Model 8, Table 3). Model 7 (where synchronous covariances were allowed to vary across time) fit the data better than did Model 8, $\chi^2_{\text{diff}}(15) = 272.20$, $p < .001$. Thus, the association between math self-concept and math grades does vary across time. Next, guided by our expectation that the association between beliefs and behaviors would be markedly higher during adolescence than during childhood, we compared a model where all synchronous covariances between Grade 1 and Grade 5 (i.e., childhood) were constrained to be equal and all synchronous covariances between Grades 6 and 12 (i.e., adolescence) were constrained to be equal (Model 9, Table 3) to the more parsimonious Model 8 (Table 3). Model 9 fit the data better than did Model 8, $\chi^2_{\text{diff}}(3) = 111.66$, $p < .001$. As expected, the estimate for the synchronous association during childhood (unstandardized = .51, standardized ~ .26) was lower than the estimate for the synchronous association during adolescence (unstandardized = 1.94, standardized ~ .51). Finally, guided by our expectation that the association between self-efficacy beliefs and behaviors would also be markedly higher during late childhood (Grades 3–5) than during early childhood (Grades 1–2), we compared a model where all synchronous covariances between Grade 1 and Grade 2 (i.e., early childhood) were constrained to be equal,

Table 3
Childhood and Beyond Model Comparisons

Model	χ^2	*df*	Chi-square, *p* value	CFI	RMSEA	χ^2_{diff}	*df* diff	*p* value
Cohort comparison								
(1) Cohort—unconstrained model	3,339.58	1425	< .001	.919	.061			
(2) Cohort—synchronous covariances constrained	3,399.95	1437	< .001	.917	.062	60.37	12	< .001
(3) Cohort—"structural" covariances constrained	3,536.77	1477	< .001	.912	.063	197.19	52	< .001
Gender comparison								
(4) Gender—unconstrained model	2,472.67	942	< .001	.932	.055			
(5) Gender—synchronous covariances constrained	2,480.63	948	< .001	.932	.055	7.96	6	.24
(6) Gender—"structural" covariances constrained	2,507.85	968	< .001	.932	.055	35.18	26	.11
Structural comparison								
(7) Unconstrained	3,339.58	1425	< .001	.919	.061			
(8) All synchronous covariances constrained	3,611.78	1440	< .001	.908	.065	272.20	15	< .001
(9) Constraining 1st–5th versus 6th–12th	3,500.13	1437	< .001	.912	.063	111.66	3	< .000[a]
(10) Constraining 1st–2nd versus 3rd–5th versus 6th–12th	3,497.86	1435	< .001	.912	.063	2.269	2	.32[b]

Note. CFI = comparative fit index; RMSEA = root mean square error of approximation.
[a] Model comparison is based on comparing Model 9–Model 8 (i.e., Model 9 fits better).
[b] Model comparison is based on comparing Model 10–Model 9 (i.e., Model 9 and 10 fit equally well).

all synchronous covariances between Grades 3 and 5 were constrained to be equal (i.e., late childhood), and all synchronous covariances between Grades 6 and 12 (i.e., adolescence) were constrained to be equal (Model 10, Table 3) to the more parsimonious Model 9 (Table 3). Model 10 did not fit the data any better than did Model 9, $\chi^2_{\text{diff}}(2) = 2.27$, $p = .32$. Thus, the synchronous correlations between math self-efficacy and math grades do not appear to differ between early childhood and late childhood.

Changes in Synchronous Correlations Between Efficacy for Aggression and Aggressive Behavior

Again, before examining whether or not the association between efficacy for aggression and aggressive behavior changed over time, we tested for cohort and sex differences in the measurement model presented in Figure 2. First, we tested for cohort differences. Specifically, a fully unconstrained model (Model 1, Table 4) where all covariances were free to vary across the two cohorts was compared to both a model where just the synchronous covariances were constrained to be equal across the two cohorts (Model 2, Table 4) and a model where all covariances were constrained to be equal across the two CDP cohorts (Model 3, Table 4). The fit of Model 1 (the fully unconstrained model) was not significantly better than either Model 2, $\chi^2_{\text{diff}}(6) = 8.53$, $p = .20$, or Model 3, $\chi^2_{\text{diff}}(26) = 35.93$, $p = .09$. Thus, we concluded that the associations between self-efficacy for aggression and aggressive behavior did not differ across the two CDP cohorts. Next, we tested for sex differences by comparing a fully unconstrained model (Model 4, Table 4) where all covariances were free to vary across sex to both a model where just the synchronous covariances were constrained to be equal across sex (Model 5, Table 4) and a model where all covariances were constrained to be equal across sex (Model 6, Table 4). The fit of Model 4 (the fully unconstrained model) did not fit the data any better than did Model 5, $\chi^2_{\text{diff}}(6) = 8.35$, $p = .21$, or Model 6, $\chi^2_{\text{diff}}(26) = 28.95$, $p = .31$. Thus, we concluded that the associations between self-efficacy of aggression and aggressive behaviors (including synchronous correlations) did not differ across sex.

Given these findings, we proceeded to estimate the simplex model of relations between self-efficacy of aggression and aggressive behaviors with the parameters constrained to be equal across sex and cohort. The resulting parameter estimates are shown in Figure 2 (note that the standardized estimates or correlations are presented opposed to the unstandardized estimates or covariances). The estimates for the synchronous correlations between self-efficacy of aggression and aggressive behaviors are of key interest. In order to test if the synchronous correlations vary over time, we conducted a series of model comparisons. First, we compared a model where all synchronous covariances were free to vary across time (Model 7, Table 4) to a model where all synchronous covariances were constrained to be equal across time (Model 8, Table 4). Model 7 (where synchronous covariances were allowed to vary across time) fit the data better than did Model 8, $\chi^2_{\text{diff}}(5) = 49.65$, $p < .001$. Thus, the association between self-efficacy of

Table 4
Child Development Project Model Comparisons

Model	χ^2	*df*	Chi-square, *p* value	CFI	RMSEA	χ^2_{diff}	*df* diff	*p* value
Cohort comparison								
(1) Cohort—unconstrained model	1,646.59	896	< .001	.899	.054			
(2) Cohort—synchronous covariances constrained	1,655.18	902	< .001	.899	.053	8.525	6	*ns*
(3) Cohort—"structural" covariances constrained	1,682.52	922	< .001	.898	.053	35.928	26	*ns*
Gender comparison								
(4) Gender—unconstrained model	1,669.37	896	< .001	.898	.054			
(5) Gender—synchronous covariances constrained	1,677.71	902	< .001	.898	.054	8.348	6	*ns*
(6) Gender—"structural" covariances constrained	1,698.31	922	< .000	.897	.054	28.945	26	*ns*
Structural comparison								
(7) Unconstrained	1,103.86	437	< .001	.909	.051			
(8) Constraining all synchronous covariances	1,153.52	442	< .001	.903	.053	49.653	5	< .001
(9) Constraining K to 3 versus 8–11	1,118.52	441	< .001	.908	.052	34.997	1	< .001[a]
(10) Constraining K to 2 versus 3 versus 8–11	1,107.35	440	< .001	.909	.051	11.167	1	< .001[b]

Note. CFI = comparative fit index; RMSEA = root mean square error of approximation.
[a] Model comparison is based on comparing Model 9–Model 8 (i.e., Model 9 fits better).
[b] Model comparison is based on comparing Model 10–Model 9 (i.e., Model 10 fits better).

aggression and aggressive behaviors does vary across time. Next, guided by our expectation that the association between beliefs and behaviors would be markedly higher during adolescence than during childhood, we compared a model where all synchronous covariances between kindergarten and Grade 3 (e.g., childhood) were constrained to be equal and all synchronous covariances between Grades 8 and 11 (i.e., adolescence) were constrained to be equal (Model 9, Table 4) to the more parsimonious Model 7 (Table 4). Model 9 fit the data better than did Model 7, $\chi^2_{\text{diff}}(1) = 34.99$, $p < .001$, demonstrating that the relations between self-efficacy of aggression and aggressive behavior were similar within childhood and within adolescence but differed between childhood and adolescence. As expected, the estimate for the synchronous association during childhood (unstandardized = .00, standardized $\sim$ .01) was lower than the estimate for the synchronous association during adolescence (unstandardized = .06, standardized $\sim$ .30). Finally, guided by our expectation that the association between beliefs and behaviors would be markedly higher during late childhood (Grades 3) than during early childhood (kindergarten—Grade2), we compared a model where all synchronous covariances during early childhood were constrained to be equal, all synchronous covariances during late childhood were constrained to be equal, and all synchronous covariances during adolescence where constrained to be equal (Model 10, Table 4) to the more parsimonious model 9 (Table 4). Model 10 fit the data better than did Model 9, $\chi^2_{\text{diff}}(1) = 11.17$, $p < .001$. As expected, the synchronous association during late childhood (unstandardized = .04, standardized $\sim$ .15) was higher than the synchronous association during early childhood (unstandardized = .00, standardized = −.02) but lower than the synchronous association during adolescence (standardized = .06, unstandardized $\sim$ .30).

Discussion

Decades of theoretical and empirical developmental research have suggested that a cognitive-developmental shift in middle childhood may account for well-documented differences between cognitive functioning in early and later childhood (Case, 1991; Harter, 1996; Sameroff & Haith, 1986). Among these developmental progressions are several that involve how children think about and judge the behaviors of self and others (Flavell & Miller, 1998; Wellman, Cross, & Watson, 2001). This article presents evidence of a similar developmental progression in links between children's self-efficacy beliefs and their own behaviors. Although many theoretical models in the social sciences imply that this link should be stable across development (Crick & Dodge, 1994; Fazio, 1995), our results suggest that, at least in the case of some self-efficacy beliefs, this assumption is not tenable.

Underlying this conclusion are our findings that the relation between beliefs in two different psychological domains (self-concept of ability for math, self-efficacy for aggression) and behaviors (grades, mother reports of child aggression) increases with development independently of the ability to measure the constructs. It is notable that these relations do not differ by sex for either math achievement or aggression where we might have expected differences to exist.

Another strong finding is that these changes replicate in four independent samples. In the case of math achievement, three independent cohorts assessed with the same self-efficacy and behavior measures across time showed the same shift in relations between childhood and adolescence and also displayed very similar relations at the same point in development (grade). The same findings replicated for self-efficacy of aggression and aggressive behaviors with two independent cohorts. Interestingly, when examining aggression beliefs and behaviors, there appears to be an additional significant shift in the relations between early childhood and middle childhood. Thus, there does appear to be an important adolescent change in the relation between self-efficacy beliefs and behaviors for both achievement and aggression beliefs but an additional shift for aggression beliefs in middle childhood. Examining whether this shift is consistent with cognitive-developmental changes that are being seen in early adolescence (Keating, 2004) will be an important avenue of research to pursue in the future.

One of the many challenges in explaining this change is accounting for the possibility that the reliability of the relevant measures may increase across time, thus producing the more stable age-related relations. Previous research has indicated that reliabilities and stability in measurement do increase by age (Davis-Kean & Sandler, 2001; Guay, Marsh, & Boivin, 2003; Trzesniewski et al., 2003), though even young children appear to respond reliably in these domains and can even make complicated differentiations across multiple aspects of the self (Marsh, Debus, & Bornholt, 2005). To rule out measurement unreliability as an alternative explanation for the present findings, our analyses corrected for attenuation by using latent variables with multiple indicators in structural equation models. Furthermore, as can be seen in the Method section, the internal reliability coefficients at both the earliest and the later ages showed adequate to good internal reliability. Similarly, the within-construct correlations across time (i.e., the extent to which beliefs and behaviors correlated with themselves across time) were both sizable and stable, revealing that belief and behavior structures remain relatively stable across time. Thus, the measures across the studies appear to be generally good indicators of the constructs measured, and the observed changes in the concurrent relations across time are unlikely to be due to changes in the reliability of the measures.

Developmental psychologists have demonstrated that children are capable of certain thoughts and actions well before the classic Piagetian stages would have predicted them (Gelman & Baillargeon, 1983). However, isolating a reliable developmental progression in belief – behavior relations implies both groups have been correct in concluding that the early development of beliefs is a precursor to later knowledge about the self. This knowledge, however, may only become accessible and predictive at a certain point in development (Harter, 1999).

The present findings thus underscore the importance of several methodological issues for future studies of belief – behavior relations. Our data rely on self-report methods of self-efficacy beliefs, and the observed changes in synchronous relations can only be generalized to other self-report findings. Similarly, beliefs other than self-efficacy beliefs should be examined to determine whether belief – behavior relations vary across types of self-beliefs. As Wellman et al. (2001) have found, false beliefs are well established by the age of 5, and thus we would not necessarily expect to see these types of changes in relations across time in tasks related to false beliefs.

Even with these limitations, the findings underscore the importance of several methodological issues for future studies of belief – behavior relations. Questions of how and when in development these phenomena should be measured also remain important for future research endeavors. Should researchers collect self-report data on self-conceptions of beliefs if the purpose is to predict behavior, especially in cross-sectional data collection? The fact that the beliefs data in the present study are self-reports from children, whereas the behavior indicators are school records and reports from parents constitutes both a strength and weakness for interpreting the analyses. Using reports from others or school records avoids the method variance problems that occur when measures of both the independent and the dependent variables come from self-reports of study participants. Experimental studies would potentially permit testing these hypotheses directly and give us additional information on the potential mechanisms underlying the change in the relations between beliefs and behaviors across time.

Despite these remaining challenges, the present evidence of changes across time in the connection between self-efficacy beliefs and behaviors is a potentially significant advance in understanding the nature of belief – behavior relations. That the finding is robust across multiple belief–behavior domains (achievement and aggression) and using different measures of behavior (grades, parent reports of aggression) strengthens the contribution of the work as a useful beginning point for understanding this phenomenon and its implications for future research on theory and methods. Finally, confirming relatively late increases in stability implies the possibility of intervening with young children at risk for potentially dysfunctional behavior-relevant beliefs. If children's beliefs regarding their academic performance or the appropriateness and effectiveness of certain behavioral strategies are in formation well into the school years, clinicians have a window of opportunity during which interventions directed toward these beliefs might lead to more positive and successful behaviors in the future.

References

Achenbach, T. M. (1991). *Child Behavior Checklist and Child Behavior Profile: Cross-Informant version*. Burlington: University of Vermont.

Alvarez, J. M., Ruble, D. N., & Bolger, N. (2001). Trait understanding or evaluative reasoning? An analysis of children's behavioral predictions. *Child Development*, *72*, 1409–1425.

Bandura, A. (1986). Social Foundation by thought and action. Englewood Clifts, NJ: Prentice Hall.

Case, R. (1991). Stages in the development of the young child's first sense of self. *Developmental Review, 11*, 210–230.

Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information processing mechanisms in children's social adjustment. *Psychological Bulletin, 115*, 74–101.

Davis-Kean, P. E., & Sandler, H. M. (2001). A meta-analysis of measures of self-esteem for young children: A framework for future measures. *Child Development*, *72*, 887–906.

Dodge, K. A., Bates, J. E., & Pettit, G. S. (1990). Mechanisms in the cycle of violence. *Science*, *250*, 1678–1683.

Eccles, J. S., Wigfield, A., Harold, R. D., & Blumenfeld, P. (1993). Ontogeny of children's self-perceptions and subjective task values across activity domains during the early elementary school years. *Child Development*, *64*, 830–847.

Fazio, R. H. (1995). Attitudes as object-evaluation associations: Determinants, consequences, and correlates of attitude accessibility. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 247–282). Hillsdale, NJ: Erlbaum.

Flavell, J. H., & Miller, P. H. (1998). Social cognition. In W. Damon (Ed.) & D. Kuhn & R. S. Siegler (Vol. Eds.), *Handbook of child psychology* (5th ed., pp. 851–898). New York: Wiley.

Gelman, R., & Baillargeon, R. (1983). A review of Piagetian concepts. In J. H. Flavell, M. E. Markman, & P. H. Mussen (Eds.), *Handbook of child psychology, Vol. 3. Cognitive development*. (pp. 167–230). New York: Wiley.

Graham, J. W., Hofer, S. M., & Piccinin, A. M. (1994). Analysis with missing data in drug prevention research. In L. M. Collins & L. A. Seitz (Eds.), *Advances in data analysis for prevention intervention research* (NIDA Research Monograph No. 142, pp. 13–63). Washington, DC: National Institute on Drug Abuse.

Guay, F., Marsh, H. W., & Boivin, M. (2003). Academic self-concept and academic achievement: Developmental perspectives on their causal ordering. *Journal of Educational Psychology, 95*, 124–136.

Harter, S. (1996). Developmental changes in self-understanding across the 5 to 7 year shift. In A. J. Sameroff & M. M. Haith (Eds.), *The five to seven year shift*. (pp. 207–236). Chicago: University of Chicago Press.

Harter, S. (1999). *The construction of the self: A developmental perspective*. New York: Guilford Press.

Huesmann, L. R. (1988). An information processing model for the development of aggression. *Aggressive Behavior, 14*, 13–24.

Huesmann, L. R., & Guerra, N. G. (1997). Children's normative beliefs about aggression and aggressive behavior. *Journal of Personality and Social Psychology, 72*, 408–419.

Joreskog, K. G. (1970). Estimation and testing of simplex models. *British Journal of Mathematical and Statistical Psychology, 23*, 121–145.

Keating, D. P. (2004). Cognitive and brain development. In R. J. Lerner & L. D. Steinberg (Eds.), *Handbook of adolescent psychology* (2nd ed., pp. 45-84). New York: Wiley and Sons.

Kline, R. B. (1998). *Principles and practice of structural equation modeling*. New York: Guilford Press.

Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In D. A. Goslin (Ed.), *Handbook of socialization theory and research*. (pp. 347–480). Skokie, IL: Rand McNally.

Lansford, J. E., Malone, P. S., Dodge, K. A., Crozier, J. C., Pettit, G. S., & Bates, J. E. (2006). A 12-year prospective study of patterns of social information processing problems and externalizing behaviors. *Journal of Abnormal Child Psychology, 34*, 709–718.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: Wiley.

Marsh, H. W. (1993). Stability of individual differences in multiwave panel studies: Comparison of simplex models and one-factor models. *Journal of Educational Measurement*, *30*, 157–183.

Marsh, H. W., Debus, R. L., & Bornholt, L. J. (2005). Validating young children's self-concept responses: Methodological ways and means to understand their responses. In D. M. Teti (Ed.), *Handbook of research methods in developmental science* (pp. 138–160). Oxford, UK: Blackwell.

Muthen, L. (2005). *Multiple imputation discussion*. Retrieved, October 15, 2007, from http://www.statmodel.com/discussion/messages/22/381.html

Muthen, L., & Muthen, B. (1998–2006). *Mplus user's guide* (4th ed.). Los Angeles: Muthen & Muthen.

Parsons, J. E., & Ruble, D. N. (1977). The development of achievement-related expectancies. *Child Development, 48,* 1075 – 1079.

Piaget, J. (2001). *The psychology of intelligence.* New York: Rutledge. (Original work published 1963)

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys.* New York: John Wiley.

Ruble, D. N., Parsons, J. E., & Ross, J. (1976). Self-evaluative responses of children in an achievement setting. *Child Development, 47,* 990 – 997.

Sameroff, A. J., & Haith, M. M. (1986). *The five to seven year shift.* Chicago: University of Chicago Press.

SAS Institute Inc. (1999). *SAS procedures guide, Version 8.* Cary, NC: Author.

Schafer, J. L. (1999). *NORM: Multiple imputation of incomplete multivariate data under a normal model. Software for Windows 95/98/NT, Version 2.*

Schafer, J., & Olsen, M. (1998). Multiple imputation for multivariate missing-data problems: A data analyst's perspective. *Multivariate Behavioral Research, 33,* 545 – 571.

Siegler, R. S., & Alibali, M. W. (2004). *Children's thinking* (4th ed.). Upper Saddle River, NJ: Prentice Hall.

Spencer, F. H., & Bornholt, L. (2003). A model of children's cognitive functioning and cognitive self concepts. *Australian Journal of Learning Disabilities, 8,* 4 – 8.

Trzesniewski, K. H., Donnellan, M. B., & Robins, R. W. (2003). Stability of self-esteem across the life span. *Journal of Personality and Social Psychology, 84,* 205 – 220.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72,* 655 – 684.

White, S. (1970). Some general outlines of the matrix of developmental changes between 5 and 7 years. *Bulletin of the Orton Society, 20,* 41 – 57.

White, S. H. (1996). The child's entry into the "age of reason." In A. J. Sameroff & M. M. Haith (Eds.), *The five to seven year shift* (pp. 17 – 30). Chicago: University of Chicago Press.

Wigfield, A., & Eccles, J. (1992). The development of achievement task values: A theoretical analysis. *Developmental Review, 12,* 265 – 310.

Child Development, September/October 2008, Volume 79, Number 5, Pages 1270–1287

Categories Influence Predictions About Individual Consistency

Marjorie Rhodes and Susan A. Gelman
University of Michigan

Predicting how people will behave in the future is a critical social-cognitive task. In four studies ($N = 150$, ages preschool to adult), young children (ages 4–5) used category information to guide their expectations about individual consistency. They predicted that psychological properties (preferences and fears) would remain consistent over time after hearing one example in which properties followed a category-linked distribution (e.g., children of different genders had different properties) but not when properties varied within a category (e.g., children of the same gender had different properties). The developmental course of these findings is examined. Results suggest the importance of considering how children's emerging theories of behavior and of social groups operate together to inform their expectations about the social world.

To an adult, hearing a child say that she "likes Winnie the Pooh" provides valuable information. For example, the adult may use this information to select birthday presents for this child or to choose books to offer to read to her. These inferences are possible because adults interpret the child's statement as stemming from a stable disposition toward Winnie the Pooh and expect this disposition to cause similar behaviors across time and situations (e.g., Yuill, 1993).

The example above demonstrates a tendency to explain human behavior by appealing to internal stable causal mechanisms, a lay theory termed *dispositionism* (Choi, Nisbett, & Norenzayan, 1999; Ross & Nisbett, 1991). Adults apply a disposition-based theory to explain and predict human behavior across a wide range of tasks, including when they have very limited information about the relevant behavior or characters (Gilbert & Malone, 1995; Uleman, 1987). A number of studies indicate, however, that preschoolers are less likely than adults to appeal to dispositions to explain behavior (Miller, 1984) or to expect individual behavior to remain stable across time and situations, especially when limited information is available (e.g., Kalish, 2002).

The goal of the present article was to examine how preschoolers approach the task of explaining and predicting behavior when they have little evidence to draw on. If young children viewed most brief observations of behavior in their everyday lives as uninformative about the future, they would have some difficulty establishing a general sense of people and their roles, which would be a considerable impediment to learning from their environment. In the present work, we examined the possibility that preschoolers rely on category knowledge to make predictions about individual future behavior. In doing so, we aimed to enrich our understanding of young children's social inferences, by examining the circumstances in which they do and do not expect people to be consistent over time, as well as to examine a previously unexplored way that categories may guide induction in early childhood.

Dispositional Thinking in Early Childhood

Previous research from the related area of trait concepts suggests that dispositional thinking is accessible to preschool children under certain circumstances (e.g., Cain, Heyman, & Walker, 1997; Heyman & Gelman, 2000) but also that young children require more (or different kinds of) evidence than adults do to engage in such reasoning (Aloise, 1993; Boseovski & Lee, 2006; Miller & Aloise, 1989). Much of the research examining dispositional thinking has tested whether children and adults use information about past experiences to guide predictions about the future. For

---

This research was supported by the National Institute of Child Health and Human Development Grant HD-36043 to the second author. The first author was supported by funding from the Michigan Prevention Research Training Grant (National Institutes of Health Grant T32 MH63057-03). We are grateful to the parents, teachers, and children of Lakewood Elementary School, Little Folks Corner, the Ann Arbor YMCA Early Childhood Center, and Concord Preschool for participating in this research, as well as to Korie Zink, Erica Ranade, and Kristin Garrison for their assistance with data collection. We also thank three anonymous reviewers who gave helpful feedback on prior drafts and whose suggestions motivated the design of Study 4.

Correspondence concerning this article should be addressed to Marjorie Rhodes, Department of Psychology, University of Michigan, 2422 East Hall, 530 Church Street, Ann Arbor, MI 48109-1043. Electronic mail may be sent to rhodesma@umich.edu.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7905-0005

example, researchers have examined whether learning that a girl liked chocolate cookies in the past leads participants to predict that she will like chocolate cookies again in the future (Kalish, 2002) or whether learning that a boy acted generously in one situation will lead participants to predict that he will act generously again in a different situation (Rholes & Ruble, 1984). On such questions, participants are often provided with relatively limited information about an actor's behavior (e.g., one example). Under these conditions, adults readily expect preferences and traits to be consistent over time, whereas children younger than age 7 demonstrate chance-level responding (Kalish, 2002; Kalish & Shiverick, 2004; Lawson & Kalish, 2006).

Using other methods, however, such as providing young children with many detailed examples of distinctive past behavior (Boseovski & Lee, 2006; Cain et al., 1997; Ferguson, van Roozendaal, & Rule, 1986; Gnepp & Chilamkurti, 1988), more sensitive response measures (Dozier, 1991), or linguistic cues (Gelman & Heyman, 1999), researchers have found that even 4-year-olds will predict behaviors and traits to be consistent over time. Also, a number of studies indicate that preschoolers can use trait terms to make predictions about behavior (Heyman & Gelman, 1999, 2000; Liu, Gelman, & Wellman, 2007) and that 3- and 4-year-olds can label examples of behavior with appropriate trait terms (Boseovski & Lee, 2006; Liu et al., 2007), indicating some facility with dispositional thinking when experimental tasks are simplified (but see Alvarez, Ruble, & Bolger, 2001; Yuill & Pearson, 1998, for alternate interpretations of preschoolers' success on tasks measuring dispositional thinking).

Kalish (2002) addressed preschoolers' hesitancy to predict consistency based on limited evidence, in spite of their knowledge of dispositions revealed on some tasks, with two proposals: (a) individuals predict consistency only when they view the event in question as resulting from a stable causal mechanism that is likely to cause the event again, and (b) preschoolers are less likely than adults to view behaviors as resulting from stable dispositions. That is, dispositions are not salient causal mechanisms within young children's theories of social behavior. Instead, Kalish suggests that young children are more likely to appeal to temporary factors to explain human behavior, such as passing desires and individual volition, and may not understand the role of stable dispositions in constraining those factors (see also Wellman, 1990). Thus, in the example given at the beginning of this article, preschoolers may conclude that the character said she liked Winnie the Pooh because she wanted to (e.g., felt like saying it), wanted Winnie the Pooh at that moment, or liked Winnie the Pooh for an external reason, such as a rule or adult request (Kalish & Shiverick, 2004; Karniol & Ross, 1976). In other words, they may not take this statement as evidence of a stable disposition toward Winnie the Pooh (see also Yuill, 1997). Because the character's preference is not viewed as resulting from a stable mechanism, preschoolers are unlikely to predict that the preference will be stable over time.

Within the framework described by Kalish (2002), it is not necessary to view dispositionism as entirely absent from young children's thinking; rather, it may be that young children require more clues before inferring the presence and importance of dispositions. For example, preschoolers may not view dispositions as relevant to as many behaviors as adults do or they may not view dispositions as common within human psychology. Thus, they would be less likely to appeal to dispositions to explain behavior if presented with little information. From this perspective, adults are more ready to predict consistency in social behavior because of the increased salience of dispositions within their naïve theories (e.g., Miller, 1984).

Category-Based Social Reasoning

We propose that when young children do not appeal to dispositions to predict behavior, they may incorporate their knowledge of other causal mechanisms into their social reasoning to make inferences. Preschoolers perceive membership in social categories, in gender categories in particular, as salient and powerful predictors of behavior. For example, Gelman, Collman, and Maccoby (1986) found that 4-year-olds expected two individuals of the same gender to share novel physical and behavioral properties, even when they were perceptually dissimilar. Similarly, Berndt and Heller (1986) found that 5-year-olds relied on gender categories to predict what sorts of activities characters would engage in (expecting them to engage in gender-typical more than gender-atypical behavior), even when they were told that characters had previously made gender-atypical behavioral choices (see also Biernat, 1991). In this case, kindergarteners relied more heavily on gender categories than adults did; adults attributed greater predictive power to characters' previous behaviors than to gender. Preschoolers also predict that children will display stereotypical gender behavior even if they are raised entirely by members of the other gender, whereas adults account for environmental influence in the development of gender-typical behavior (Taylor, 1996; Taylor, Rhodes, & Gelman, in press).

As discussed in detail by Lawson and Kalish (2006), preschoolers' willingness to base inferences on single pieces of evidence on category-based induction tasks, such as those involving gender, is in stark contrast to their hesitancy to use a single piece of evidence to make an inference about an individual over time. Preschoolers appear to consider membership in certain categories as predictive of a wide range of properties, such that they expect members of the same category to share many nonobvious properties, even if they differ from each other in other ways, such as physical appearance (for a review, see Gelman, 2003). Further, preschoolers expect membership in these informative categories to remain stable over time, regardless of environment (Gelman & Wellman, 1991; Hirschfeld, 1996; Taylor, 1996). Thus, whereas preschoolers may appeal to temporary factors to explain instances of individual behavior, they appeal to a stable factor (category membership) to make predictions about properties across category members.

We propose that in addition to appealing to category-based mechanisms to predict consistency across category members, preschoolers attend to social cues that guide them to appeal to categories to understand and predict instances of individual behavior as well. In situations where preschoolers do not perceive dispositions as causing behavior, they may attend to evidence indicating that the behavior is caused by membership in an important social category (e.g., gender). If such evidence is found, we expect preschoolers to infer that the behavior is caused by membership in a gender category, and therefore, to predict that the behavior will be consistent over time. Because gender categories are salient to young children, we expect that they will conclude that gender is causally relevant to behavior based on very limited information, as well as when the relation between gender and behavior is only implicitly communicated to them. Specifically, we hypothesize that preschoolers will predict consistency in individual behavior when they learn that individuals of one gender category, and not the other, display the properties of interest.

The Present Studies

In these studies, we first developed questions about characters' preferences and choices, in which participants received limited information about characters' previous experience. All questions involved characters engaging in novel activities, such that children's prior beliefs about the particular activities would not influence their pattern of responses. These questions did not elicit disposition-based reasoning among preschoolers, consistent with prior research. Next, we presented children with questions about pairs of characters who each experienced two things and had unique reactions to them. For example, a pair of children played two games; one child liked one game, whereas the other child liked the other. In some questions, the pairs involved two children of different genders, such that a boy liked one game and a girl liked the other. In this case, we expected children to perceive the pattern as indicating that liking a particular game is causally related to gender and, therefore, to expect preferences to be stable over time. The options were never labeled as "for boys" or "for girls"; rather, children were given information about the preferences of individual male and female characters and were left to draw their own conclusions about the relation between gender and behavior. In other questions, participants were told about pairs of children of the same gender who each liked different activities. In this case, because the pattern does not imply that liking a particular activity is causally related to gender, we did not expect preschoolers to predict consistency over time.

Developmental Changes in the Use of a Category-Based Strategy

We also included older comparison groups in these experiments, in order to examine how the use of a category-based strategy to predict individual consistency changes over time. Because adults require less information than preschool children do to make dispositional inferences (Aloise, 1993), rely on a category-based strategy as often. On questions in which adults view characters' behaviors as relevant to dispositions, they should predict consistency regardless of category information.

Adults do not engage in dispositional thinking on all questions about individual behavior, however. For example, Kalish (2002) found that on questions asking whether characters would make consistent *choices* between two options that differed rather trivially (e.g., a boy who chose a blue flower over a yellow flower), adults responded randomly. This is likely because multiple strategies seem appropriate for answering this question; the boy might choose blue because of a stable disposition toward blue or he might choose yellow because he already has a blue flower (Kalish, 2002). Because no information about preferences is provided in these questions, both responses seem plausible.

We hypothesized that when adults perceive some ambiguity as to whether an actor's behavior should be considered as caused by a disposition, they too will turn to information about category membership to guide their predictions about consistency. We also hypothesized, however, that adults will appeal to category memberships only when they view the relevant categories as likely to have caused the property of interest. As reviewed above, adults view the influence of gender as less global and less deterministic than do younger children. Although adults often hold some strong gender stereotypes, in general they view gender as influencing a relatively narrower range of behaviors and view causal forces aside from gender (e.g., the environment, personality) as important predictors of behavior (Taylor, 1996). Therefore, given the paucity of information about the relation between gender and behavior presented in these questions, as well as that questions involve *novel* (i.e., nonstereotyped) behaviors, we expected that adults would be less likely than preschoolers to assume that characters' choices in these stories were caused by gender.

We included comparison questions about nonhuman animal species in order to determine whether any age-related changes in the use of a category-based strategy related to a complete shift away from viewing category-based mechanisms as relevant to individual behaviors or to developmental changes in the meaning of gender per se. These questions followed a parallel structure to all our questions about humans and gender, including questions about individual animals (e.g., one dog), pairs of same-species animals (e.g., two dogs), and pairs of different-species animals (e.g., a dog and a cat). We selected nonhuman animals because there is evidence that both young children and adults view animal species as indicating deep and fundamental differences between individuals and view many properties as causally related to membership in an animal kind (Gelman, 2003; Medin & Ortony, 1989). Therefore, we expected both preschoolers and adults to rely on species information to predict consistency (on questions in which they do not rely on dispositions).

Thus, young children were expected to reason similarly about gender categories and animal species, treating both kinds of categories as strongly influencing behavior, whereas adults were expected to show more differentiated reasoning about gender and animal species (Taylor, 1996; Taylor et al., in press). In later experiments, we also included a comparison sample of 10-year-olds to examine more precisely the development of differentiated reasoning about gender and species categories on these questions. Because we used novel properties in all questions, we were able to compare reasoning about humans and nonhuman animals directly. This research will therefore contribute to our understanding of how young children use category information to make predictions about individual consistency as well as of the influence of different kinds of categories on reasoning across development.

Overview

The aim of Study 1 was to develop a series of questions about single individuals and their psychological properties to be used in our subsequent experiments. In Study 2, we added information about category membership (of gender and animal categories) to the questions, such that preference for a particular activity either did or did not vary according to category membership. In Study 3, we expanded the type of properties and age ranges included. Finally, in Study 4, we demonstrated that preschoolers' predictions are influenced by gender categories but not by more superficial differences (i.e., shirt color).

Study 1

The goal of Study 1 was to develop questions to be used in our subsequent experiments. Participants were told about single individuals (animals or humans) who experienced two activities, had a particular response to them, and then a few days later were presented with the same two options again. We purposefully provided participants with limited information about prior behavior so that in subsequent experiments we could assess whether preschoolers engage in category-based reasoning to approach these questions. We included two types of psychological properties, preferences and choices, which we expected to differ in the extent to which they cued adults to appeal to dispositions. Thus, we hypothesized that preschoolers would demonstrate chance-level responding on all questions, whereas adults would predict consistency on questions about preferences but not on questions about choices.

Method

Participants. Participants (n = 30; 13 male and 17 female) included 14 preschoolers (M age = 4 years 8 months, range = 4 years 0 months – 5 years 7 months) recruited from private preschools and 16 adults (M age = 19 years 2 months, range = 17 years 8 months – 22 years 1 month) recruited from introductory psychology classes at a large Midwestern public university. In all studies, the samples were approximately 80% White, 15% Asian American, and 5% from other racial/ethnic backgrounds. All participants were tested individually in a quiet office (located either at the children's schools or an on-campus laboratory) by the first author or by trained female undergraduate research assistants who were blind to study hypotheses. College students received course credit for their participation.

Procedure. Participants were presented with 16 questions, including 8 questions about children (4 about boys and 4 about girls) and 8 questions about animals of various species. Half of the questions asked about characters' preferences and half asked about their choices. Thus, there were 4 questions of each of four types (e.g., human preferences, human choices, animal preferences, and animal choices). All questions followed an identical format (see Appendix A).

For example, on a question about preferences, participants were told, "Here is Tom. A few days ago, Tom did two things: Tom went flecking and Tom went tooping. Tom liked flecking. Today, Tom is doing two things again. Again, Tom is going flecking and tooping. What will Tom like today? Will Tom like flecking, like last time, or will Tom like tooping?" After one question of each type, randomly selected in advance, participants were asked, "Remind me, which did [Tom] like the first time?" This allowed us to assess whether participants were following the scenarios; no adult and only 1 child had difficulty answering any of these questions. This child missed only one control question (of four total asked) and so was retained for analyses. All questions were accompanied by simple pictures of the story character as well as by small cards with the names of the activities printed on them. Pointing was used to aid comprehension. For all questions, responses in which participants predicted that the character would act in a consistent manner were scored as 1 (the alternate response was scored as 0).

Questions were blocked by character type (animal or human child), and order of presentation of the two blocks was counterbalanced across participants. Within each block, questions were presented in a separate random order for each participant. Several precautions were taken to prevent the results from being biased by any unforeseen preferences that the participants might have had for particular nonsense words. Specifically, the order of presentation of the two activities and the selection of particular novel activities by the story character were counterbalanced for each question across participants (e.g., in the example question above, half of the participants were told that Tom liked "flecking," whereas the other half were told that he liked "tooping"). Also, pairs of nonsense words circulated throughout the question types across participants, such that each pair of words was used equally in each of the four types of questions.

Results

We first compared the proportion of questions on which participants predicted that characters would behave consistently to the proportion expected by chance (chance = .5) for each type of question, using a series of sign tests. As shown in Figure 1, on questions about preferences, preschoolers demonstrated chance-level responding (human children: *M* = 0.48, *SD* = 0.37; animals: *M* = 0.44, *SD* = 0.65), whereas adults predicted consistency more often than expected by chance (human children: *M* = 0.78, *SD* = 0.22; animals: *M* = 0.83, *SD* = 0.22). On questions about choices, preschoolers again did not predict consistency; they demonstrated chance-level responding on questions about animals (*M* = 0.37, *SD* = 0.42) and predicted consistency *less* often than expected by chance on questions about human children (*M* = 0.25, *SD* = 0.35). As hypothesized, adults also did not reliably predict consistency on questions about choices (human children: *M* = 0.41, *SD* = 0.31; animals:

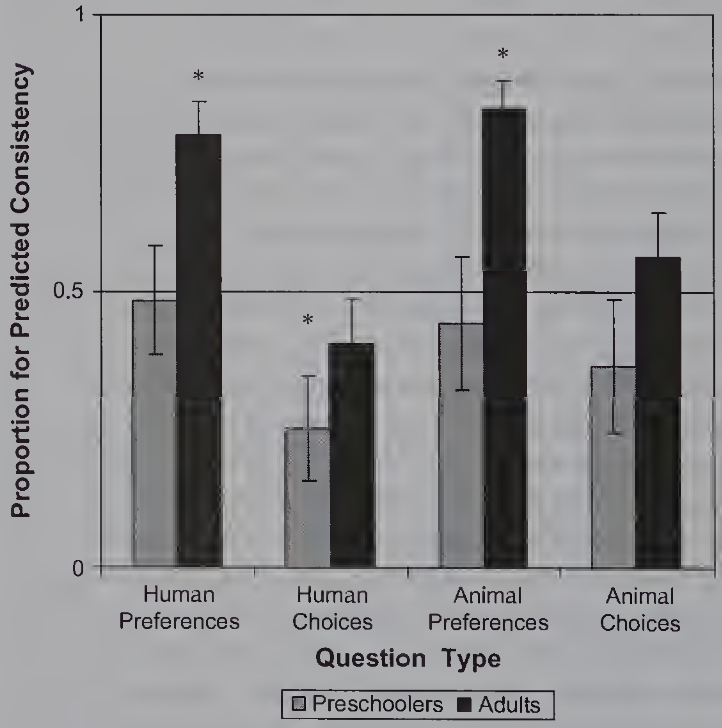


Figure 1. Proportion for predicted consistency by age and question type, Study 1.
Note. Error bars represent standard errors of the mean. Sign tests comparing the obtained proportions of predicted consistency to the proportion expected by chance.
$^{*}p < .05$.

$M = 0.56$, $SD = 0.34$). There were no main or interactive effects of participant gender or character gender, so these variables were excluded from analyses.

We next examined the data through a 2 (age: preschooler, adult) × 2 (character: human, animal) × 2 (question type: preference, choice) repeated measures analysis of variance, (ANOVA), with character and question type as within-subjects variables and age as a between-subjects variable. This analysis yielded main effects for age, $F(1, 27) = 6.12$, $p < .05$, such that adults predicted consistency more often than did preschoolers, and for question type, $F(1, 27) = 24.83$, $p < .001$, such that consistency was predicted more often for preferences than for choices. There were no other significant main or interactive effects. The interaction between question type and age was marginal, $F(1, 27) = 3.06$, $p = .09$; however, follow-up analyses indicated that the effect of age was specific to questions about preferences. As demonstrated in Figure 1, adults predicted consistency more often than preschoolers on questions about preferences, for humans, $t(28) = 2.71$, $p < .05$, $d = .99$; for animals, $t(28) = 3.0$, $p < .01$, $d = 1.10$, but not on questions about choices, for humans, $t(28) = 1.28$, *ns*; for animals, $t(28) = 1.41$, *ns*.

Discussion

The results from Study 1 support the hypothesis that preschoolers are less likely than adults to predict that preferences are consistent over time; adults strongly expected preferences to be consistent over time, whereas children responded at chance. Also, as predicted, adults did not reliably predict consistency on questions about choices. Interestingly, on the questions about human choices presented in this study as well as on some similar questions about choices asked by Kalish (2002), young children actually chose the alternate (inconsistent) response more often than expected by chance. This suggests that children may expect others to favor variety *over* consistency, especially when no preference information is provided.

Although these findings regarding children's beliefs about choices are interesting, in Study 2, we decided to ask children questions about preferences only, for two reasons. First, the aim of this series of studies was to examine whether children would use category information to guide their predictions about consistency on questions on which they otherwise would not respond in an adult-like manner, and adults reliably expected consistency only on questions about preferences. Second, pilot testing demonstrated that because the questions asked in Study 2 contained more information than those from Study 1 (see below), the task was too long for young children when questions about both preferences and choices were asked.

Based on these considerations, both preschoolers and adults in Study 2 were asked questions about preferences. For adults only, we also included supplemental questions about choices. These were included because adults' predictions of consistency on questions about preferences were relatively high, indicating that they were engaging in dispositional thinking on these questions. Therefore, we did not expect information about category memberships to influence their predictions. Results from Study 1 indicated, however, that they did not engage in dispositional thinking on questions about choices; therefore, adding category information to these questions allowed us to test the hypothesis that adults are influenced by information about animal species, and not gender categories, on these questions.

Study 2

In Study 2, we tested whether guiding participants to view individuals' psychological properties as linked to category membership would lead them to predict greater levels of individual consistency. Participants were asked questions about pairs of either animals or human children that followed a similar format and referred to the same activities as those asked in Study 1. In Study 2, however, participants were given scenarios about two characters in which one character preferred one activity and the other character preferred the other activity. We hypothesized that when the pairs consisted of two individuals that shared category membership (e.g., two children of the same gender or two animals of the same species), the property would not be viewed as linked to category membership (as two members of the same category have two different preferences), and therefore, young children would again fail to predict consistency. In contrast, when participants are told about pairs consisting of members of different categories (e.g., two children of different genders or two animals of different species), we predicted that children would then view the property as category linked and would predict greater levels of individual consistency. We hypothesized that adults would readily predict consistency for all questions about preferences, regardless of category information. On the supplemental questions about choices, however, we hypothesized that they would be influenced by information about pair composition on questions about animal species but not human gender.

Method

Participants. Participants (n = 53, 23 male and 30 female) included 24 preschoolers (*M* age = 4 years 7 months; range = 3 years 11 months–5 years 8 months) and 29 adults (*M* age = 19 years 1 month; range = 17 years 9 months–22 years 7 months) drawn from the same population as those who participated in Study 1; no one participated in both studies.

Procedure. All participants were tested individually in a quiet office (located at either the children's schools or an on-campus laboratory) by the first author or by trained female undergraduate research assistants who were blind to study hypotheses. College students received course credit for their participation.

To limit the number of questions that each child was asked, participants were randomly assigned to receive questions about either animals or humans. Each condition included eight questions, which consisted of two blocks of four questions each, one block asked about pairs of individuals from the same category (i.e., same gender or same species) and the other block asked about pairs of individuals from different categories (i.e., different gender or different species). Order of presentation of the two blocks was counterbalanced across participants, and the questions within each block were asked in a separate random order for each participant. For adults, each block contained an additional four questions describing choices instead of preferences (see Appendix B). The same procedures as in Study 1 were used for assigning nonsense words to questions for each participant and for counterbalancing the order of nonsense words within each pair.

Questions followed the same structure and used the same properties as in Study 1, except that pairs of characters were described and none were given individual names (see Appendix B). For example, children heard: "Here are two children, a boy and a girl. A few days ago these children did two things, they went lexing and they went gorping. The boy liked lexing. The girl liked gorping. Today, the boy is doing those two things again. Again, the boy is going lexing and gorping. What will the boy like today? Will the boy like lexing, like last time, or will the boy like gorping?" After one question of each type, randomly selected in advance, participants were asked to recall the details of the story after they made their predictions, with the following questions, "Remind me … which one did [the boy] like the first time? And, which one did [the girl] like?" No participant had difficulty answering these control questions. Questions were accompanied by pictures of both characters and cards with the names of the activities printed on them. Pointing was used to aid comprehension. After the first set of preferences was described, the picture of the nontarget character was removed.

Results

We first present the analyses examining children's and adults' predictions about preferences, followed by supplemental analyses examining adults' predictions about choices. Preliminary analyses revealed that there were no main or interactive effects of participant gender or character gender; therefore, these variables were not included in analyses.

As shown in Figure 2, sign tests revealed that preschoolers predicted consistency at levels exceeding those expected by chance for questions about pairs of individuals from different categories (different gender: *M* = 0.71, *SD* = 0.32; different species: *M* = 0.85, *SD* = 0.27). In contrast, preschoolers predicted consistency less often than expected by chance on questions about same-gender pairs (*M* = 0.27, *SD* = 0.33) and at chance levels for same-species animal pairs (*M* = 0.60, *SD* = 0.34). Adults predicted consistency more often than expected by chance for all questions about preferences (same gender: *M* = 0.69, *SD* = 0.28; different gender: *M* = 0.75, *SD* = 0.27; same species: *M* = 0.83, *SD* = 0.28; different species: *M* = 0.89, *SD* = 0.23).

We further examined these data through a 2 (age: preschooler, adult) × 2 (condition: human

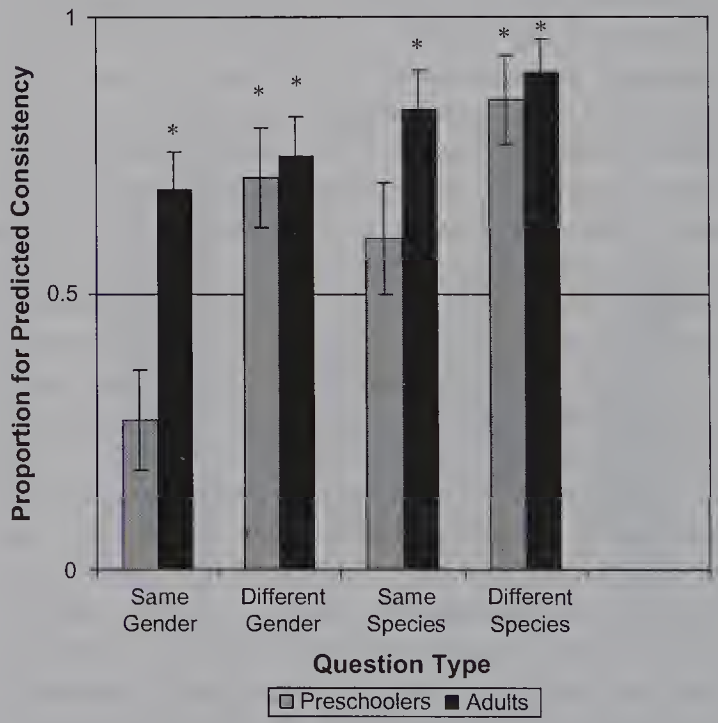


Figure 2. Proportion for predicted consistency by age and pair composition, Study 2.
Note. Error bars represent standard errors of the mean. Sign tests comparing the obtained proportions of predicted consistency to the proportion expected by chance.
$^*p < .05$.

children, animals) × 2 (pair composition: same category, different category) repeated measures ANOVA, with age and condition as between-subjects factors and pair composition as a within-subject factor. This analysis revealed main effects for pair composition, $F(1, 49) = 18.15$, $p < .001$, such that more consistency was predicted on questions about members of different categories; condition, $F(1, 49) = 8.86$, $p < .01$, such that more consistency was predicted for animals than people; and age, $F(1, 49) = 8.29$, $p < .001$, such that adults predicted consistency more often than did children. There was also a significant Pair Composition × Age interaction, $F(1, 49) = 8.48$, $p < .01$; see Figure 2. Planned contrasts revealed that preschoolers predicted consistency more often on questions involving pairs from different categories than on questions involving pairs from the same category ($p < .001$, $d = 1.20$), but adults' predictions did not differ depending on the composition of the pairs ($p > .5$). This pattern was identical and significant for questions about human and animal targets. Also, for same-category pairs, preschoolers predicted consistency less often than adults did ($p < .001$, $d = -1.04$), but for different-category pairs, they predicted consistency as often as did adults ($p > .3$).

Adults' predictions about choices. As hypothesized, adults predicted consistency more often than expected by chance only for questions describing animals of different species ($M = 0.75$, $SD = 0.28$, $p < .01$, sign test) and not for any other question type (same species: $M = 0.50$, $SD = 0.45$; different gender: $M = 0.45$, $SD = 0.26$; same gender: $M = 0.45$, $SD = 0.31$). Also as hypothesized, adults' predictions on questions about animals differed depending on the composition of the pair; adults predicted consistency more often on items describing different-species animals than on items describing same-species animals, $t(14) = 2.49$, $p < .05$, $d = .66$. In contrast, adults' predictions on questions about humans did not differ depending on the gender composition of the pair, $t(15) < .001$, *ns*

Discussion

In Study 2, we found support for the hypothesis that guiding preschoolers toward viewing individuals' preferences as linked to category membership leads them to predict greater individual consistency. Children predicted that preferences would remain consistent after learning that members of different categories had different preferences but not after learning that members of the same category had different preferences. We also documented that whereas adults' predictions about properties for which they have strong expectations of consistency (e.g., preferences) are not influenced by category information, they do attend to category information on questions about properties for which their expectations of consistency are less certain. Adults' concepts of different kinds of categories also appear to influence whether they view a category as relevant to individual consistency on these questions; adults were influenced only by information about animal species not by information about gender categories.

The results from adults' predictions about choices provide preliminary support for our hypothesis that categories are only used to predict individual consistency when the category is viewed as causally relevant to novel behaviors. We have not yet been able to document age-related changes in the use of gender and animal species on questions about the same psychological property, however, because adults predict preferences to be consistent so readily, and young children in this study were not asked questions about choices. We address this limitation in Study 3 by including a comparison group of 10-year-olds. We selected 10-year-olds for two reasons. First, previous research indicates that 10-year-olds predict consistency more often than do younger children, but we did not expect their predictions to be as stable as those of adults; therefore, we expected their predictions to be somewhat influenced by information about categories. Second, previous research indicates that by age 10, children have substantially refined their concepts of gender and view the influence of gender as relatively probabilistic and specific, as opposed to as deterministic and global (e.g., Berndt & Heller, 1986; Taylor, 1996; Taylor et al., in press). Therefore, we expected their inferences—like those of adults—to be influenced by information about animal species but not human gender. Comparing the responses of 10-year-olds with those of younger children will allow for a cleaner test of the hypothesis that beliefs about particular categories influence whether children appeal to category-based mechanisms.

Study 3

In Study 3, we further examined the developmental course of the results presented in Studies 1 and 2 by including a group of 10-year-olds. To maximize comparability between the two age groups, our younger participants in this study were drawn from kindergarten classes from the same elementary school attended by the 10-year-olds, who were drawn from fifth-grade classrooms. Thus, the participants in the younger age group of this study were slightly older than in Studies 1 and 2 but still within the age range of young children who are unlikely to engage in dispositional thinking on questions that provide limited evidence (e.g., Kalish, 2002).

We also made two small modifications designed to address the generality of our findings. First, we provided participants with information about what kinds of activities the characters engaged in. In Studies 1 and 2, participants were told relatively little information, for example, that characters had "gone lexing." They were left to infer that "lexing" is some sort of activity, but it could be a game, a chore, or any other type of activity. In Study 3, we examined whether children would use the same category-based strategies as they had in Study 2 when information about the types of activities that characters engaged in is explicitly given. For example, in these questions, children are told about boys and girls who played two games or bears and monkeys that climbed two kinds of trees. The options were still labeled using nonsense words (e.g., to refer the particular games or trees) in order to control for the influence of children's prior knowledge or experiences (e.g., their preferences for specific games). For example, children were told that characters played the "lex game and the gorp game."

Second, in Study 3, we expanded the types of psychological properties examined. In Study 2, children were asked questions about preferences only. Because preferences are an area that children frequently think about in terms of gender (e.g., games that girls like), it is possible that children's gender-based strategy is limited to questions about preferences. Alternately, because young children view gender as having a broad influence on psychological properties, they may apply this strategy more liberally. To distinguish between these possibilities, in Study 3, we included questions about both preferences and fears. We selected fears because older children tend to view fears as stable over time, whereas younger children may not (e.g., Kalish, 2002; Yuill, 1997).

Because Study 3 involved new sets of questions and new ages, we first conducted an initial pilot study following the same format as in Study 1. The goal of this pilot study was to confirm that 10-year-olds predicted consistency more often than 5-year-olds on questions about single individuals, using the new properties created for this study, when no information about categories is provided. After the pilot study demonstrated the expected age-related differences on our new items, we conducted the experimental study, in which participants were asked a series of questions about pairs of children and animals (similar to Study 2).

Method

Participants. Participants ($n = 55$; 26 male and 29 female) included 25 kindergarteners (M age = 5 years 7 months, range = 4 years 9 months–6 years 5 months) and 30 fifth graders (M age = 10 years 1 months, range = 9 years 2 months–11 years 4 months) recruited from a single public elementary school in a Midwestern university town. Eleven of the kindergartners and 15 of the fifth graders were assigned to the pilot study, and the rest were assigned to the main study (described below).

Procedures. Participants were tested in a quiet office at their elementary school by either the first author or trained female undergraduate research assistants who were blind to study hypotheses. In both the pilot study and the main study, participants were asked a total of 16 questions (8 about animal pairs and 8 about human pairs). Questions were blocked by character type, and order of presentation of these blocks was counterbalanced across participants; kindergarteners completed the blocks in two sessions, approximately 1 week apart; fifth graders completed the study in a single session. Within each block, questions were also blocked by pair type (same category or different category), with order of these blocks also counterbalanced across participants. Within each block, questions were presented in a separate random order for each participant. Procedures for counterbalancing nonsense words were the same as in previous studies.

Pilot study. In the pilot study, participants were asked questions about individual actors—eight human characters and eight animal characters—which followed the same structure as the questions used in Study 1 but referred to the new properties developed for this study. Half of the questions asked about preferences and half asked about fears, such that there were four questions of each of four types (four animal preferences, four animal fears, four human preferences, and four human fears; see Appendix C). Results from this pilot study indicated that the questions using the new properties elicited the expected age-related findings. Specifically, kindergarteners did not predict consistency more often than expected by chance on any questions (human children: $M = 0.36$, $SD = 0.44$; animals: $M = 0.30$, $SD = 0.43$; $p > .34$, sign tests), whereas fifth graders predicted consistency more often than expected by chance for both kinds of characters (human children: $M = 0.70$, $SD = 0.33$; animals: $M = 0.70$, $SD = 0.31$; $p < .05$, sign tests). Also, fifth graders predicted consistency more often than did kindergarteners on questions about both humans, $t(24) = 2.22$, $p < .05$, $d = .87$, and animals, $t(24) = 2.78$, $p < .05$,

d = 1.08. There were no effects associated with property type (preference or fear).

The questions asked in this pilot study were similar in design and purpose to those asked in Study 1, with three changes: (a) the properties were modified as described above, (b) the younger participants were slightly older than in Study 1, and (c) category labels were used to refer to the characters in the question (e.g., "This is a girl") instead of proper names (e.g., "This is Molly"). Despite these changes, the proportion of questions on which kindergarteners predicted consistency did not differ significantly from the proportion of questions on which preschoolers in Study 1 predicted consistency, on questions about human children, $t(23) = .73$, *ns*, or animals, $t(23) = .82$, *ns*. Based on these findings, we developed the questions for the main study using the new property descriptions.

Main study. In the main study, participants were asked eight questions about animals (four same-species pairs and four different-species pairs) and eight questions about children (four same-gender pairs and four different-gender pairs). Thus, there were four questions of each of four types (see Appendix D). Procedures were similar to Study 2 and materials included the same pictures of the story characters, as well as cards containing the names of the target activities.

Results

In the main study, we first compared the proportion of questions on which participants predicted that characters would behave consistently to the number expected by chance for each type of question, followed by repeated measures ANOVAs. Preliminary analyses revealed that there were no main or interactive effects of participant gender or character gender; therefore, this variable was not considered in analyses. Also, there were no differences between preferences and fears for either age group; in no analysis did the main or interactive effect of property type (preference or fear) approach significance (all $ps > .4$). Therefore, we collapsed across this distinction.

As shown in Figure 3, sign tests revealed that kindergarteners predicted consistency more often than expected by chance on items involving different category members for questions about both human gender ($M = 0.80$, $SD = 0.33$) and animals ($M = 0.73$, $SD = 0.40$) but not for questions involving same-category pairs (same gender: $M = 0.41$, $SD = 0.35$; same species: $M = 0.43$, $SD = 0.36$). Fifth graders predicted consistency more often than expected by chance for items involving different animal species only ($M = 0.75$, $SD = 0.27$) and not for items involving same-species animals ($M = 0.63$, $SD = 0.39$), same-gender children ($M = 0.68$, $SD = 0.35$), or different-gender children ($M = 0.62$, $SD = 0.38$).

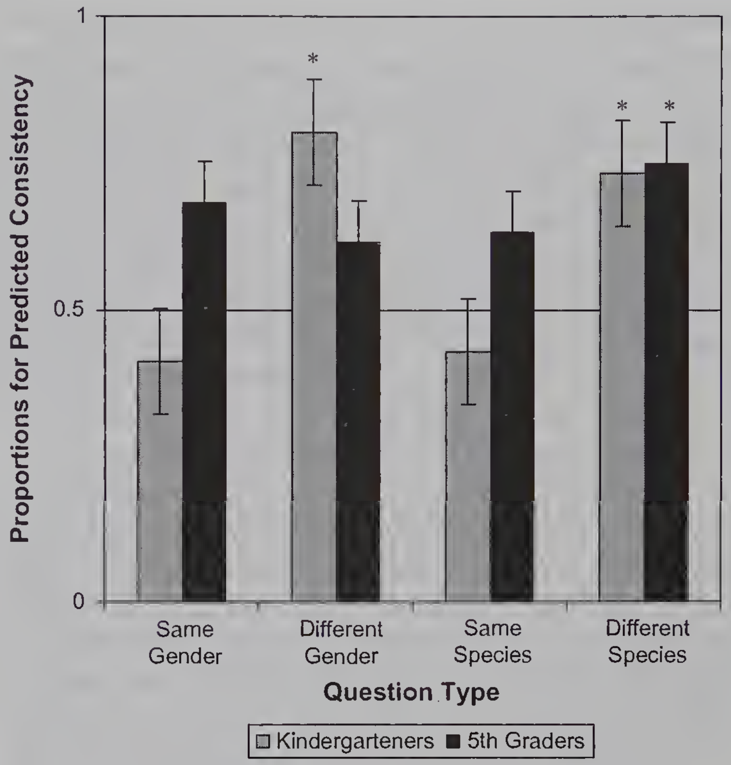


Figure 3. Proportion for predicted consistency by age and pair composition, Study 3.
Note. Error bars represent standard errors of the mean. Sign tests comparing the obtained proportions of predicted consistency to the proportion expected by chance.
$*p < .05$.

We further examined these data through a 2 (target: animal, human) × 2 (pair composition: same category, different category) × 2 (age: kindergarten, fifth grade) repeated measures ANOVA, with target and pair as within-subjects factors and age as a between-subjects factor (see Figure 3). This analysis revealed a main effect of pair composition, $F(1, 27) = 16.64$, $p < .001$, and a significant interaction between pair composition and age, $F(1, 27) = 12.48$, $p < .01$. Planned contrasts revealed that kindergartners predicted consistency more often on questions about different-category members than on questions about same-category members, for both human and animal characters ($p < .001$; gender, $d = 1.15$; animal, $d = .79$). Although the three-way interaction among target, pair composition, and age failed to reach significance ($p = .1$), for fifth graders, as hypothesized, pair composition did not influence responses on items about gender: Predictions about same-gender and different-gender pairs did not differ ($p > .3$). However, pair composition did influence responses on items about animals: Predictions of consistency were made significantly more often about different-species pairs than about same-species pairs, $t(14) = 2.43$, $p < .05$, $d = .35$.

Discussion

In Study 3, we replicated the findings from Studies 1 and 2 for younger children and extended these findings to questions about fears. When presented with questions involving individual human children or animals, kindergartners did not expect the characters to have consistent preferences or fears over time (as demonstrated in the pilot study). When presented with pairs of children or animals, however, kindergartners expected characters to have consistent preferences and fears only when they were told that members of different categories had different properties.

In contrast, fifth graders generally expected humans and animals to have consistent preferences and fears, as evidenced in the pilot study. Also as hypothesized, fifth graders predicted consistency more often on questions about different-species animals than on questions about same-species animals but did not differentiate between questions about same-gender and different-gender pairs. These findings are consistent with prior work demonstrating that older children view gender as less influential for behavior than do younger children (e.g., Berndt & Heller, 1986; Taylor, 1996). These findings, as well as the findings from Study 2, suggest that developmental changes in how gender influences predictions of consistency on these questions relate to changes in children's gender concepts not to a general move away from viewing category information as relevant to individual consistency.

Surprisingly, fifth graders demonstrated chance-level responding on all questions in the main study, except for questions involving pairs of different animals. Thus, learning that pairs of children had different preferences and fears may have led fifth graders away from predicting consistency. Although the predictions of fifth graders did not differ from chance in the main study, they did still predict consistency more often than kindergarteners. Nevertheless, it is somewhat surprising that fifth graders' predictions on these questions were not above chance, as they are well within the age range of children that have been shown to engage in dispositional reasoning in other studies (e.g., Kalish, 2002).

Children's spontaneous explanations that accompanied these questions provide some insight into this unexpected pattern. Participants frequently mentioned various social learning mechanisms in explaining their responses. For example, children suggested that characters might "learn to like it from their friend" or "become afraid of it because their partner did" (explanations fitting this interpretation were given at least once by 8 of 14 fifth graders in this condition). Thus, children hypothesized that social interactions between the characters might influence the tendency to try new things or remain consistent. Importantly, this process was not influenced by the gender composition of the pair, supporting the hypothesis that fifth graders would not view gender as influencing the consistency of these psychological properties. It is also possible, given children's explanations, that their beliefs about the social context presented in the study questions somehow guided them away from viewing gender as relevant. Species information, however, did influence this pattern, as described above. Thus, although children posited social interactions that would lead the pairs of children and animals to influence each other, they did not expect such interactions to operate across species boundaries (although they did expect them to operate across gender categories). Overall, these findings support the prediction that, with age, children develop more differentiated concepts of animal species and gender categories.

Study 4

Study 4 was a control study, which was designed to test whether young children appeal to category-based mechanisms only when the category information that is presented refers to meaningful differences between people. In Studies 2 and 3, young children reliably predicted consistency on questions involving both different-gender humans and different-species animals. One possibility is that young children will appeal to *any* observable difference between actors to make predictions about consistency, even if the distinction is relatively unlikely to influence the behavior of interest. From this perspective, the more children are guided to focus on the differences between two characters, the more likely they may be to infer stable individual differences. If this were the case, then young children's responses on our previous studies would not be indicative of their beliefs about the causal power of gender or species categories per se and would not represent a theory-driven category-based approach to social reasoning. Alternately, if children use gender categories, but not relatively more trivial differences between people, to predict consistency, then the pattern of findings documented in our earlier studies may be interpreted as reflecting the conceptual status of gender categories and animal species. To distinguish between these two possibilities, we compared children's reasoning about pairs of different-gender children and pairs of children who differ on another easily observable (but less meaningful) dimension, in this case, shirt color. We selected shirt color because it is easy for young children to track visually, amenable to labeling, and has been used in other studies designed to compare reasoning about deep and relatively superficial categories (Solomon, 2002).

Method

Participants. Participants included 12 preschoolers (5 male and 7 female; *M* age = 4 years 8 months, range = 4 years 1 month–5 years 7 months) recruited from the same population areas as children who participated in Studies 1 and 2. None had participated in an earlier study.

Procedure. Children were tested individually, following the procedures used in previous studies. All children were asked a series of eight questions: four about different-gender pairs and four about pairs of children (of the same gender) who wore different-color shirts. Items were blocked by pair type (gender or shirt color), and order of presentation of the blocks was counterbalanced across participants. Within the blocks, items were presented in a separate random order for each participant. The questions about gender began, "See these two children? This one is a boy. This one is a girl," and then were identical to the different-gender pair questions from Study 2. The questions about shirt color began "See these two children? This one is wearing a yellow shirt. This one is wearing a green shirt," and then followed the same structure as the other questions (see Appendix E). As in previous studies, children were presented with simple pictures of the story characters, as well as with small cards with the names of the nonsense activities printed on them, and pointing was used to aid comprehension. New pictures were used for this experiment, so that we could control shirt color (for different-gender pairs, the children wore the same-color shirt; for different-shirt-color pairs, children were of the same gender, but one wore a yellow shirt and one wore a green shirt). As in previous studies, order of presentation of each nonsense word within a pair and the character's selected preference were counterbalanced across participants and the pairs of nonsense words circulated across items.

Results

Consistent with previous studies, preschoolers predicted that characters' preferences would remain consistent more often than expected by chance when they were told about different-gender pairs ($M = 0.73$, $SD = 0.31$, $p < .05$, sign test). As hypothesized, however, they did not predict that preferences would be consistent when they were told about pairs of children wearing different-color shirts; in fact, they predicted that preferences would be consistent marginally less often than expected by chance ($M = 0.27$, $SD = 0.39$, $p = .09$, sign test), which was similar to preschoolers' performance on questions about same-gender pairs in Study 2. Children predicted consistency more often on questions about gender pairs than on questions about shirt-color pairs, $t(11) = 5.01$, $p < .001$, $d = 1.53$.

Discussion

In this control study, we documented that preschoolers apply a category-based strategy only when the individual differences that are presented relate to meaningful categories (i.e., gender) but not when they relate to more trivial differences (i.e., shirt color). These findings support our interpretation that individuals appeal only to categories that they view as having an important causal influence on behavior to predict consistency and suggest that preschoolers appeal to categories in a strategic manner to predict the future.

General Discussion

We examined whether individuals rely on category information to predict individual behavior. We found that preschoolers, older children, and adults all rely on such a process to guide their social reasoning. We also found, however, that there are age differences in how commonly participants engaged in category-based reasoning, as well as in which kinds of categories individuals found informative for reasoning about these items.

We found robust evidence for a category-based approach to reasoning about individual consistency among preschoolers and kindergarteners. When young children were told that a single individual experienced two options and had a preference for or was afraid of one of them, they were no more likely to predict that those responses would be consistent over time than expected by chance, congruent with prior work (e.g., Kalish, 2002). When young children were told that two characters from different salient categories had unique responses (e.g., a boy liked one option, whereas a girl liked the other), however, they predicted that psychological responses would be consistent over time more often than expected by chance and as often as adults did. We found evidence for increased predictions of consistency only when the two characters differed in terms of membership in an important category; young children did not predict consistency when they were told about the unique responses of two characters that shared category membership (e.g., two girls), even if the two characters differed in some other observable manner (e.g., shirt color).

Young Children's Category-Based Reasoning

Why do preschoolers predict consistency when they learn that members of different categories have different properties? It is well documented that young children view membership in certain categories as fundamental to identity and as determining a wide range of behaviors (Gelman, 2003). Much of this research has focused on children's beliefs about animal species and has demonstrated that young children make a broad range of inferences about animals' properties based on species information alone (Gelman & Markman, 1986; Gelman & Wellman, 1991; Sousa, Atran, & Medin, 2002; Waxman, Medin, & Ross, 2007). As described by Gelman, (2003), species categories take on this inductive potential because children view category membership as tied to the presence of a category "essence," or a nonobvious internal entity that exists in every category member and causes category membership and the development of category-linked properties. Thus, the development of category properties is viewed as causally related to category membership.

A number of studies indicate that children also perceive membership in some social categories as resulting from the presence of a category essence (Astuti, Solomon, & Carey, 2004; Diesendruck & haLevi, 2006; Hirschfeld, 1996; Taylor, 1996). Preschoolers treat gender categories in particular as having high inductive potential and as causally determining behavior. For example, preschoolers expect members of the same-gender category to share novel properties even in the face of contradictory perceptual or environmental information (Gelman et al., 1986; Taylor, 1996). Thus, preschoolers appear to use gender categories and animal species information in a similar manner in inductive reasoning, treating them both as indicating membership in natural kinds (Rothbart & Taylor, 1992).

Information about category membership, whether regarding animal species or human gender, presents young children with powerful causal explanations for behavior. When confronted with the question of whether a particular behavior is likely to be consistent over time, the belief that a behavior is causally related to category membership provides a strong reason to predict consistency. We propose that because categories provide such important explanatory information, young children are sensitive to detecting category-linked distributions. In the present studies, they formed expectations that a behavior was tied to gender after a single piece of evidence. Behaviors were never labeled as "for boys" or "for girls," children made this inference on the basis of information about only two individuals.

Developmental Changes in Category-Based Reasoning

Adults and older children (fifth grade) also relied on information about categories to guide their inferences about consistency, but only on questions involving animal species, and for adults, only on questions about particular psychological properties. We will discuss two aspects of the development of social cognition that we believe account for these age-related findings, one related to the increased salience of dispositionism as a lay theory for explaining behavior, and the other related to changing beliefs about gender.

Dispositionism. Kalish (2002) suggests that dispositions, or internal stable individual differences, are not salient causal mechanisms within young children's theories of behavior. Although previous work suggests that preschoolers have some capacity to engage in dispositional thinking (e.g., Heyman & Gelman, 2000), on problems about individual consistency they appear to require more or different kinds of evidence than older children and adults before they will view a given behavior as good evidence on which to base an inference of future consistency.

This interpretation is consistent with findings from cross-cultural research indicating increasing cultural divergence in social attribution with age. These cross-cultural findings suggest that social attribution is, at least in part, the result of an extended process of socialization (Miller, 1984, 1986). Wellman and Miller (2006) summarize this body of literature and suggest that although there are some cross-cultural commonalities in young children's understanding of people and behavior (e.g., viewing people as active and intentional agents), the specifics of children's naïve theories of psychology, including which factors they view as salient and powerful behavioral motivators, vary cross culturally. Among adults as well, some aspects of social attribution appear to be common across cultures (e.g., viewing individuals as active in initiating their own behavior), but individuals in different cultural contexts appear to differ in the extent to which they also attribute a potent causal role to the environment (Choi et al., 1999; Morris & Peng, 1994; Norenzayan, Choi, & Nisbett, 2002). Thus, attributing a primary causal role to individual dispositions appears to be the result of a process of socialization, interacting with a probably universal tendency to view humans as intentional. From this perspective, it is not expected that preschoolers would share the same conceptual framework for explaining and predicting behavior as the adults in their environment, as much of socialization is yet to be completed (Wellman & Miller, 2006). Thus, the increasing salience of dispositionism with age in the present context may contribute to our finding that older children and adults relied on a category-based strategy less often than did younger children.

Gender concepts. A second developmental process that may contribute to the present findings relates to developmental changes in concepts of different kinds of categories. Young children reasoned about animal species and human gender in a similar manner, regardless of whether this variable was manipulated within or between participants, suggesting that young children attribute a similar causal role to both kinds of categories. Young children's relatively strict concepts of gender have been suggested to relate to a general tendency to focus on within-category similarity and between-category differences and to ignore within-category variability (Gelman, 2003; Gelman & Coley, 1991; Maccoby, 1988). Their approach to gender may also relate to specific beliefs about *how* gender influences behavior. For example, previous work suggests that preschoolers view engaging in gender-stereotyped behaviors as somewhat automatic, or as necessary for fulfilling moral or social obligations, as opposed to as subject to environmental or intentional control (Levy, Taylor, & Gelman, 1995; Taylor et al., in press).

In contrast, fifth graders' and adults' reasoning appeared to reflect more differentiated concepts of animal species and human gender; they were influenced by category information about animal species only. Prior work has suggested that, over time, individuals come to view gender as influencing a narrower set of properties, as well as to view the influence of gender on behavior as probabilistic and dependent on environmental pressure and personal agency (Berndt & Heller, 1986; Taylor, 1996). Although adults and older children often hold some strict stereotypes about gender, these may be more specific than the theories held by young children. With age, individuals' concepts of gender focus on differences in particular areas, whereas younger children expect more general differences (Maccoby, 1988). Thus, we propose that developmental changes in theories about gender also contribute to our age-related findings.

We have indicated that because young children view gender categories as such strong predictors of behavior, they are willing to infer a gender-linked cause for behavior after one example of a gender-linked pattern. In contrast, the presented evidence was not sufficient to warrant such an interpretation by older children or adults. We suspect, however, that older children and adults would rely on gender if they were presented with a greater amount of gender-linked information (e.g., questions involving stereotypical properties or more examples of gender-linked patterns). In future research, it would be interesting to compare how much gender-linked information preschoolers, older children, and adults require before they view novel behaviors as related to gender.

Limitations and Future Directions

A direction for future research suggested by these findings stems from the fact that the information about past experiences provided to participants in the present experiments was relatively minimal, as was the information provided about categories. In the present work, we found that when both of these sources of information are minimal (e.g., one example of a prior behavior and one example of a category-linked distribution), preschoolers relied on category-linked information but not dispositions to guide their inferences. One limitation of this approach, however, is that all the activities referred to in the present research were novel. Although using novel activities in this context was useful for experimental control, it is certainly possible that preschoolers will infer dispositions based on single acts if they are able to apply their knowledge about particular activities (e.g., they may view some activities, such as liking brussel sprouts, as distinct enough as to warrant dispositional reasoning).

Also, because the information provided was limited, we cannot determine from the present research when children might rely on a category-based strategy in more realistic contexts. Based on prior work, we suspect that if more information about previous behavior was provided, preschoolers may be more likely to view the behavior as linked to a disposition (e.g., Boseovski & Lee, 2006). Providing more information about gender, however, should also lead to heightened reliance on gender categories to predict behavior. In future research, it would be interesting to examine how young children integrate information about past experiences and social categories when both sources of information are highly salient. Using realistic examples of behavior, Berndt and Heller (1986) found that young children preferred to base inferences about the future on gender stereotypes rather than on past behavior; however, children in this study were also given relatively limited information (see also Biernat, 1991). Thus, this remains an important area for future work.

These findings are also limited by the cultural context in which we conducted this research. Specifically, the age-related decline in the influence of gender-linked information documented in the present studies must be interpreted within the context of the present sample. All the data for these experiments were collected in a relatively liberal university town. Thus, it is possible that other cultural contexts support the maintenance of strict gender concepts throughout development. Interestingly, even in the relatively liberal cultural context sampled in the present work, we still found that young children hold relatively strong views regarding the influence of gender on behavior and reasoned about gender categories in a comparable manner to how they reasoned about animal species. In less egalitarian contexts, older children and even adults may continue to rely on gender to guide inferences about individual consistency when limited evidence is available. We interpret the present findings as indicating that whether individuals rely on a category-based strategy when making social inferences depends on the causal power attributed to the category for determining a wide range of behaviors. Whether gender is viewed in this manner should depend on age (as documented in the present work) as well as context (e.g., Crouter, Whiteman, & McHale, 2007).

In future research, it would also be interesting to examine how individuals reason about other social categories. We demonstrated that preschoolers appeal to membership in gender categories, but not more trivial differences based on shirt color, to guide their inferences. Based on our framework, we would expect that other social categories that individuals perceive as causal motivators of behavior to be used in a similar manner. For example, prior work suggests that young children attribute causal influence to categories based on race (Hirschfeld, 1996) and ethnicity (Astuti et al., 2004; Diesendruck & haLevi, 2006). Thus, we would expect children (and adults) to rely on these categories and others as well, to the extent that they view the categories as having a strong influence on a wide range of known and unknown behaviors.

In sum, these results inform two aspects of the development of social cognition, children's understanding of the mechanisms leading to individual consistency and the ways in which concepts of social categories influence children's expectations about the social world. These findings extend knowledge of how young children make sense of the social world, by pointing toward cues that preschoolers use when determining which behaviors are likely to be engaged in consistently over time. The present findings also document the importance of considering how young children's emerging theories of behavior and of social groups operate together to inform their explanations and predictions about the social world.

References

Aloise, P. A. (1993). Trait confirmation and disconfirmation: The development of attribution biases. *Journal of Experimental Child Psychology, 55*, 177–193.

Alvarez, J. M., Ruble, D. N., & Bolger, N. B. (2001). Trait understanding or evaluative reasoning? An analysis of children's behavioral predictions. *Child Development, 72*, 1409–1425.

Astuti, R., Solomon, G. E. A., & Carey, S. (2004). Constraints on conceptual development. *Monographs of the Society for Research in Child Development*, 69, (2, Serial No. 277).

Berndt, T. J., & Heller, K. A. (1986). Gender stereotypes and social inferences: A developmental study. *Journal of Personality and Social Psychology, 50*, 889–898.

Biernat, M. (1991). Gender stereotypes and the relationship between masculinity and femininity: A developmental analysis. *Journal of Personality and Social Psychology, 61*, 351–365.

Boseovski, J. J., & Lee, K. (2006). Children's use of frequency information for trait categorization and behavioral prediction. *Developmental Psychology, 42*, 500–513.

Cain, K. M., Heyman, G. D., & Walker, M. E. (1997). Preschoolers' ability to make dispositional predictions within and across domains. *Social Development, 61*, 53–75.

Choi, I., Nisbett, R. E., & Norenzayan, A. (1999). Causal attribution across cultures: Variation and universality. *Psychological Bulletin, 125*, 47–63.

Crouter, A. C., Whiteman, S. D., & McHale, S. M. (2007). Development of gender attitude traditionality across middle childhood and adolescence. *Child Development, 78*, 911–926.

Diesendruck, G., & haLevi, H. (2006). The role of language, appearance, and culture in children's social category-based induction. *Child Development, 77*, 539–553.

Dozier, M. (1991). Functional measurement assessment of young children's ability to predict future behavior. *Child Development, 62*, 1091–1099.

Ferguson, T. J., van Roozendaal, J., & Rule, B. G. (1986). Informational basis for children's impressions of others. *Developmental Psychology, 22*, 335–341.

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday life.* New York: Oxford University Press.

Gelman, S. A., & Coley, J. D. (1991). Language and categorization: The acquisition of natural kind terms. In S. A. Gelman & J. P. Byrnes (Eds.), *Perspectives on language and thought: Interrelations in development* (pp. 146–196). New York: Cambridge University Press.

Gelman, S. A., Collman, P., & Maccoby, E. E. (1986). Inferring properties from categories versus inferring categories from properties: The case of gender. *Child Development*, *57*, 396–404.

Gelman, S. A., & Heyman, G. D. (1999). Carrot-eaters and creature-believers: The effects of lexicalization on children's inferences about social categories. *Psychological Science*, *10*, 489–493.

Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition*, *23*, 183–209.

Gelman, S. A., & Wellman, H. M. (1991). Insides and essences: Early understandings of the nonobvious. *Cognition*, *23*, 183–209.

Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin*, *117*, 21–38.

Gnepp, J., & Chilamkurti, C. (1988). Children's use of personality attributions to predict other people's emotional and behavioral reactions. *Child Development*, *59*, 743–754.

Heyman, G. D., & Gelman, S. A. (1999). The use of trait labels in making psychological inferences. *Child Development*, *70*, 604–619.

Heyman, G. D., & Gelman, S. A. (2000). Preschool children's use of trait labels to make inductive inferences. *Journal of Experimental Child Psychology*, *77*, 1–19.

Hirschfeld, L. A. (1996). *Race in the making: Cognition, culture, and the child's construction of human kinds*. Cambridge, MA: MIT Press.

Kalish, C. W. (2002). Children's predictions of consistency in people's actions. *Cognition*, *84*, 237–265.

Kalish, C. W., & Shiverick, S. M. (2004). Children's reasoning about norms and traits as motives for behavior. *Cognitive Development*, *19*, 401–416.

Karniol, R., & Ross, M. (1976). The development of causal attributions in social perceptions. *Journal of Personality and Social Psychology*, *34*, 455–464.

Lawson, C. A., & Kalish, C. W. (2006). Inductive inferences across time and identity: Are category members more alike than single individuals? *Journal of Cognition and Development*, *7*, 233–252.

Levy, G. D., Taylor, M. G., & Gelman, S. A. (1995). Traditional and evaluative aspects of flexibility in gender roles, social conventions, moral rules, and physical laws. *Child Development*, *66*, 515–531.

Liu, D., Gelman, S. A., & Wellman, H. M. (2007). Components of young children's trait understanding: Behavior-to-trait inferences and trait-to-behavior predictions. *Child Development*, *78*, 1543–1558.

Maccoby, E. E. (1988). Gender as a social category. *Developmental Psychology*, *24*, 755–765.

Medin, S., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179–195). New York: Cambridge University Press.

Miller, J. G. (1984). Culture and the development of everyday social explanation. *Journal of Personality and Social Psychology*, *46*, 961–978.

Miller, J. G. (1986). Early cross-cultural commonalities in social explanation. *Developmental Psychology*, *22*, 514–520.

Miller, P. H., & Aloise, P. A. (1989). Young children's understanding of the psychological causes of behavior: A review. *Child Development*, *60*, 257–285.

Morris, M. W., & Peng, K. (1994). Culture and cause: American and Chinese attributions for social and physical events. *Journal of Personality and Social Psychology*, *67*, 949–971.

Norenzayan, A., Choi, I., & Nisbett, R. E. (2002). Cultural similarities and differences in social inference: Evidence from behavioral predictions and lay theories of behavior. *Personality and Social Psychology Bulletin*, *28*, 109–120.

Rholes, W. S., & Ruble, D. N. (1984). Children's understanding of dispositional characteristics of others. *Child Development*, *55*, 550–560.

Ross, L., & Nisbett, R. E. (1991). *The person and the situation: Perspectives of social psychology*. New York: Mcgraw-Hill.

Rothbart, M., & Taylor, M. (1992). Category labels and social reality: Do we view social categories as natural kinds? In G. R. Semin & K. Fiedler (Eds.), *Language, interaction and social cognition* (pp. 11–36). Thousand Oaks, CA: Sage Publications.

Solomon, G. A. (2002). Birth, kind, and naïve biology. *Developmental Science*, *5*, 213–218.

Sousa, P., Atran, S., & Medin, D. (2002). Essentialism and folkbiology: Evidence from Brazil. *Journal of Cognition and Culture*, *2*, 195–223.

Taylor, M. (1996). The development of children's beliefs about social and biological aspects of gender differences. *Child Development*, *67*, 1555–1571.

Taylor, M., Rhodes, M., & Gelman, S. A. (in press). Boys will be boys; cows will be cows: Children's essentialist reasoning about human and animal development. *Child Development*.

Uleman, J. S. (1987). Consciousness and control: The case of spontaneous trait inferences. *Personality and Social Psychology Bulletin*, *12*, 337–354.

Waxman, S., Medin, D., & Ross, N. (2007). Folkbiological reasoning from a cross-cultural developmental perspective: Early essentialist notions are shaped by cultural beliefs. *Developmental Psychology*, *43*, 294–308.

Wellman, H. M. (1990). *The child's theory of mind*. Cambridge, MA: MIT Press.

Wellman, H. M., & Miller, J. G. (2006). Developing conceptions of responsive intentional agents. *Journal of Cognition and Culture*, *6*, 27–55.

Yuill, N. (1993). Understanding of personality and dispositions. In M. Bennett (Ed.), *The development of social cognition: The child as psychologist* (pp. 87–110). New York: Guilford Press.

Yuill, N. (1997). English children as personality theorists: Accounts of the modifiability, development, and origins of traits. *Genetic, Social, and General Psychology Monographs*, *123*, 5–26.

Yuill, N., & Pearson, A. (1998). The development of bases for trait attribution: Children's understanding of traits as causal mechanisms based on desire. *Developmental Psychology*, *34*, 574–586.

Appendix A

Sample Items Study 1

Human preference	Here is Megan. A few days ago, Megan did two things—Megan went famming and Megan went lopping. Megan liked famming. Today, Megan is doing two things again. Again, Megan is going famming and lopping. What will Megan like today? Will Megan like famming, like last time, or will Megan like lopping?
Human choice	Here is Billy. A few days ago, Billy had a choice. Billy could either go rilling or Billy could go taffing. Billy went rilling. Today, Billy has a choice again. Again, Billy can either go rilling or taffing. What will Billy do today? Will Billy go rilling, like last time, or will Billy go taffing?
Animal preference	Here is Buttercup. A few days ago, Buttercup did two things. Buttercup went soting and Buttercup went tabing. Buttercup liked soting. Today, Buttercup is doing two things again. Again, Buttercup is going soting and tabing. What will Buttercup like today? Will Buttercup like soting, like last time, or will Buttercup like tabing?
Animal choice	Here is Torty. A few days ago Torty had a choice. Torty could either go vanning or Torty could go jerning. Torty went vanning. Today Torty has a choice again. Again, Torty can either go vanning or jerning. What will Torty do today? Will Torty go vanning, like last time, or will Torty go jerning?
Nonsense words	Pipping, famming, lexing, zazzing, lopping, gorping, tooping, fepping, jerning, flecking, daxing, vanning, spouding, soting, heesting, bemming, tabbing, sugging, pagging, sudding, shaining, flooming, molling, risting, fulfing, thaying, blicking, plizzing, rilling, hepping, taffing, pomming

Appendix B

Sample Items Study 2

Same-gender pair	Here are two boys. A few days ago, they did two things; they went bemming and they went spouding. This boy liked bemming. This boy liked spouding. Today, this boy is doing two things again. Again, this boy is going bemming and spouding. What will this boy like today? Will this boy like bemming, like last time, or will this boy like spouding?
Different-gender pair	Here are two children, a girl and a boy. A few days ago they did two things; they went shaining and they went risting. The girl liked shaining. The boy liked risting. Today, the girl is doing two things again. Again, the girl is going shaining and risting. What will the girl like today? Will the girl like shaining, like last time, or will the girl like risting?
Same-species pair	Here are two pigs. A few days ago, they did two things; they went flooming and they went pagging. This pig liked flooming. This pig liked pagging. Today, this pig is doing two things again. Again, this pig is going flooming and pagging. What will this pig like today? Will this pig like flooming, like last time, or will this pig like pagging?
Different-species pair	Here are two animals, a lion and a bear. A few days ago they did two things, they went thaying and they went plizzing. The lion liked thaying. The bear liked plizzing. Today, the lion is doing two things again. Again, the lion is going thaying and plizzing. What will the lion like today? Will the lion like thaying, like last time, or will the lion like plizzing?
Sample choice question (asked to adults only)	Here are two children, a girl and a boy. A few days ago they had a choice. They could either go shaining, or they could go risting. The girl went shaining. The boy went risting. Today, the girl has a choice again. Again, the girl can either go shaining or risting. What will the girl do today? Will the girl go shaining, like last time, or will the girl go risting?

Appendix C
Sample Items Study 3, Pilot Study

Human preference	Here is a boy. A few days ago, this boy had two things for a snack, a fep and a dax. This boy liked the fep. Today, this boy is having two things for a snack again. Again, this boy is having a fep and a dax. What will this boy like today? Will this boy like the fep, like last time, or will this boy like the dax?
Human fear	Here is a girl. A few days ago, this girl saw two animals, a toop and a fleck. This girl was afraid of the toop. Today, this girl sees two animals again. Again, this girl sees a toop and a fleck. What will this girl be afraid of today? Will this girl be afraid of the toop, like last time, or will this girl be afraid of the fleck?
Animal preference	Here is a monkey. A few days ago, this monkey climbed two kinds of trees, floom trees and pag trees. This monkey liked climbing floom trees. Today, this monkey is climbing trees again. Again, this monkey is climbing floom trees and pag trees. What will this monkey like climbing today? Will this monkey like floom trees, like last time, or will this monkey like pag trees?
Animal fear	Here is a mouse. A few days ago, this mouse saw two animals, a lex and a gorp. This mouse was afraid of the lex. Today, this mouse sees two animals again. Again, this mouse sees a lex and a gorp. What will this mouse be afraid of today? Will this mouse be afraid of the lex, like last time, or will this mouse be afraid of the gorp?

Appendix D
Sample Items Study 3, Main Study

Same-gender pair *(preference)*	Here are two boys. A few days ago, these boys played two games—the fulf game and the blick game. This boy liked the fulf game. This boy liked the blick game. Today, this boy is playing two games again. Again, this boy is playing the fulf game and the blick game. What will this boy like today? Will this boy like the fulf game, like last time, or will this boy like the blick game?
Different-gender pair *(fear)*	Here are two children, a girl and a boy. A few days ago, these children saw two animals, a toop and a fleck. The girl was afraid of the toop. The boy was afraid of the fleck. Today, the girl sees two animals again. Again, the girl sees a toop and a fleck. What will the girl be afraid of today? Will the girl be afraid of the toop, like last time, or will the girl be afraid of the fleck?
Same-species pair *(fear)*	Here are two turtles. A few days ago, they saw two animals, a pom and a hep. This turtle was afraid of the pom. This turtle was afraid of the hep. Today, this turtle sees two animals again. Again, this turtle sees a pom and a hep. What will this turtle be afraid of today? Will this turtle be afraid of the pom, like last time, or will this turtle be afraid of the hep?
Different-species pair *(preference)*	Here are two animals, a bear and a monkey. A few days ago, they climbed two kinds of trees, rist trees and shain trees. The bear liked climbing rist trees. The monkey liked climbing shain trees. Today, the bear is climbing trees again. Again, the bear is climbing rist trees and shain trees. What will the bear like climbing today? Will the bear like climbing rist trees, like last time, or will the bear like climbing shain trees?

Appendix E
Sample Items Study 4

Different-gender pair	See these two children? This one is a boy. This one is a girl. A few days ago, they did two things; they went bemming and they went spouding. The boy liked bemming. The girl liked spouding. Today, the boy is doing two things again. Again, the boy is going bemming and spouding. What will the boy like today? Bemming, like last time, or spouding?
Different-shirt-color pair	See these two children? This one is wearing a green shirt. This one is wearing a yellow shirt. A few days ago they did two things; they went shaining and they went risting. The one wearing a green shirt liked shaining. The one wearing a yellow shirt liked risting. Today, the one wearing a green shirt is doing two things again. Again, the one wearing a green shirt is going shaining and risting. What will the one wearing a green shirt like today? Shaining, like last time, or risting?

Child Development, September/October 2008, Volume 79, Number 5, Pages 1288 – 1309

Stability and Change of Moral Disengagement and Its Impact on Aggression and Violence in Late Adolescence

Marinella Paciello, Roberta Fida, Carlo Tramontano, Catia Lupinetti, and
Gian Vittorio Caprara
Sapienza University of Rome

Stability and change of moral disengagement were examined in a sample of 366 adolescents from ages 14 to 20 years. Four developmental trajectories were identified: (a) nondisengaged group that started with initially low levels followed by an important decline, (b) normative group that started with initially moderate levels followed by a decline, (c) later desister group that started with initially high-medium levels followed by an increase from 14 to 16 years and an even steeper decline from 16 to 20 years, and (d) chronic group that started with and maintained medium-high levels. The results attest that adolescents who maintained higher levels of moral disengagement were more likely to show frequent aggressive and violent acts in late adolescence.

Among the mechanisms conducive to aggression, moral standards exert a notable influence in dictating when it may be legitimate to resort to behavior that may hurt other people and when, on the contrary, one has to refrain from detrimental conduct (Bandura, 1986, 1991, 2001). People are active agents who pursue their goals in accordance with personal standards that serve as guides and deterrents for action through self-reactive anticipatory evaluations. People monitor their conduct and the conditions under which it occurs, judging their actions in relation to their moral standards and perceived circumstances, and regulate their actions by anticipating the consequences they would apply to themselves. People do things that give them satisfaction and refrain from behaviors that bring self-censure. Anticipatory self-pride and self-blame are suggested to be human mind self-regulatory capacities that keep behavior in line with personal standards. Yet, this does not exclude the possibility that people, after having adopted personal standards and despite being morally committed to ethical principles, may enact behaviors that violate those standards, while continuing to profess those principles and avoiding feelings of conflict, guilt, or remorse. The chronicles of recent atrocities attest to the fact that even considerate people may become engaged in inhumane behavior without experiencing any moral concern or discomfort (Bandura, 1999, 2004; Zimbardo, 1995, 2004).

In this regard, Bandura's (1986) social cognitive theory provides a theory of *moral agency* that supplies the conceptual apparatus that is needed to clarify the mechanisms by which people come to live in accordance with moral standards as well as the mechanisms that allow a kind of divorce between moral thought and moral actions (Bandura, 1991; Bandura, Barbaranelli, Caprara, & Pastorelli, 1996). In particular, Bandura (1986, 1991) has introduced the construct of moral disengagement to explain the determinants and mechanisms governing aggressive behaviors. Moral disengagement refers to individual's tendency to use mechanisms conducive to a selective disengagement of moral censure. This is achieved by reconstructing behavior, obscuring causal agency, misrepresenting injurious consequences, and blaming victims. These mechanisms allow individuals to engage in self-serving behavior that is in contrast with their moral principles, while continuing to advocate those principles and without incurring self-evaluative emotional reactions such as guilt. It is likely that these mechanisms become crystallized over time when dealing with transgressions in the pursuit of self-interest. Indeed, a previous study found that moral disengagement to be positively correlated with self-enhancement and negatively correlated with self-transcendence (Caprara & Capanna, 2005).

The focus of the present study is to examine the stability and change of moral disengagement both at

---

This study was partially supported by grants from the Spencer Foundation and W. T. Grant Foundation to Albert Bandura, and from the Johann Jacobs Foundation and Ministero dell'Istruzione dell'Univesità e della Ricerca to G.V.C. (COFIN 1998, 2000, 2004) and to Eugenia Scabini (COFIN 2000 – 2002). We are grateful to the anonymous reviewers and Marie S. Tisak for their helpful comments and suggestions.

Correspondence concerning this article should be addressed to Marinella Paciello, Department of Psychology, Sapienza University of Rome, Via dei Marsi 78, 00185 Rome, Italy. Electronic mail may be sent to marinella.paciello@uniroma1.it.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7905-0006

a population level and at an interindividual level as well as explore the extent to which different developmental pathways are associated with different aggressive and violent outcomes over the course of adolescence. Over the past few decades, different cognitive theories have examined moral development and the relation among moral thinking, social behaviors, and social context (Bandura, 1986; Crick & Dodge, 1994; Dodge & Crick, 1990; Gibbs, Potter, Barriga, & Liau, 1996; Huesmann, 1988, 1998; Kohlberg, 1969; Turiel, 1978, 1983). Furthermore, extensive knowledge has been accumulated on the development of processes related to moral self-regulation, such as the construction of moral standards, the formation of moral judgments, and the exercise of self-influence through emotional self-reactions (Bandura, 1991; Eisenberg, 2000; Killen & Smetana, 2006; Tangney, Stuewig, & Mashek, 2007; Turiel, 2006). According to social cognitive theory and adopting a developmental approach, we focus on self-exonerative mechanisms that exert a crucial role in attenuating the links between moral thought and moral actions. These mechanisms have been overlooked in the examination of moral development. To our knowledge, this is the first study that longitudinally maps individual differences in moral disengagement over the course of adolescence.

The present study focused on the adolescent transition because various manifestations of aggression may dramatically change throughout childhood, adolescence, and early adulthood, and these changes are not the same for all individuals (Brame, Nagin, & Tremblay, 2001; Loeber & Stouthamer-Loeber, 1998; Moffitt, 1993; Sampson & Laub, 2003; Tremblay, 2000). Most adolescents adopt antisocial behavior patterns but a large part abandon them in adulthood and only a small percentage become deeply and chronically engaged in risky aggressive and violent behaviors (Loeber, 1991; Loeber & Hay, 1997; Tremblay, 2000). We believe that moral disengagement can be crucial in explaining mechanisms conducive to chronic engagement in aggressive and violent behavior (Bandura, 1999; Bandura, Caprara, Barbaranelli, Pastorelli, & Regalia, 2001; Bandura et al., 1996; Zimbardo, 2004, 2006).

Moral Disengagement

Over the course of development, the adoption of moral standards attests that the property of human mind is guided from within in accordance with principles that derive from one's own and others' experiences and that, in various ways, are at the core of social life. Individuals' increasing self-regulatory capacities and the change from external to autonomy regulation make adolescence a period particularly sensitive for the study of moral functioning. During adolescence, the development of metacognitive abilities promotes the internalization of moral principles, and ultimately, the development of one's own moral identity and agency (Bergman, 2002; Blasi, 1984; Carlo, Eisenberg, & Knight, 1992; Hardy & Carlo, 2005; Helwig & Turiel, 2003; Kohlberg & Candee, 1984; Krettenauer, 2004; Swanson & Hill, 1993). Once formed, moral standards dictate the goals to be pursued and the actions from which to refrain and exert their influence over behavior through anticipated positive and negative self-evaluative reactions. Although self-sanctions carrying feelings of guilt are associated with actions that violate individual moral standards, positive reactions carrying pride, and self-worth are associated with actions that promote those moral standards (Bandura, 1991; Bandura & Walters, 1959; Elkin & Westley, 1955; Emmons & Diener, 1986). Self-sanctions exert an important influence in refraining individual's detrimental conduct particularly when one's interest is at stake, and detrimental conduct may be instrumental in the pursuit of specific goals. However, the adoption of moral abstract principles, although necessary, is not always sufficient to refrain from detrimental conduct. Through moral disengagement, self-sanctions can be deactivated and affective self-evaluative reactions are avoided, permitting different types of detrimental conduct while saving the same moral standards.

In the social cognitive theory of moral agency, there are four major points at which self-sanction can be disengaged from detrimental conduct and eight mechanisms that have been corroborated experimentally, through which moral disengagement operates (Bandura, 1991). Different disengagement practices may focus on the *behavior*, on the sense of *personal responsibility*, on the *outcomes* of behavior, and on the *recipients* of behavior. Thus, the first set of disengagement practices—focusing on behaviors—operates on the cognitive construal of the behavior itself. It includes *moral justification*, *euphemistic language*, and *advantageous comparison*. By moral justification, detrimental conduct is made personally and socially acceptable (Kelman & Hamilton, 1989; Rapoport & Alexander, 1982; Reich, 1990; Sanford & Comstock, 1971). A lot of aggressive behavior gets justified in the name of protecting honor and reputation (Cohen & Nisbett, 1994). Euphemistic language provides a convenient way of masking immoral activities conferring a respectable status upon them (Bollinger, 1982; Diener, Dineen, Endresen, Beaman, & Fraser, 1975; Lutz, 1987). People behave much more aggressively when assaulting a person is given a sanitized label than when it is called aggression. By advantageous comparison, detrimental conduct can lose its repugnancy by comparing it with more flagrant inhumanities (Bandura, 1991). The more flagrant the contrasted activities, the more likely it is that one's own injurious conduct will appear trifling or even benevolent. The second set of disengagement mechanisms—focusing on personal responsibility—operates by obscuring, minimizing, or disclaiming the agentic role in the harm that one causes. It includes *displacement of responsibility* and *diffusion of responsibility* by which people view their actions as ordered by the social pressures of others rather than as something for which they are personally responsible (Bandura, Underwood, & Fromson, 1975; Diener, 1977; Milgram, 1974; Zimbardo, 1995). People are willing to behave in ways they normally repudiate if a legitimate authority accepts responsibility for the effects of their actions and if responsibility can be diffused (when everyone is responsible, no one really feels responsible). The third set of disengagement mechanisms—focusing on outcomes—avoids self-deterring reactions by *disregarding* or *distorting the consequences of action*. When people are involved in activities harmful to others for personal gain or because of social inducements, they avoid facing the harm they cause or they minimize it (Klass, 1978). They readily recall prior information given to them about the potential benefits of the behavior, but are less able to remember its harmful effects (Brock & Buss, 1962, 1964). In addition to selective inattention and cognitive distortion of effects, the misrepresentation may involve active efforts to discredit evidence of the harm they cause. The final set of disengagement practices focusing on the recipients of detrimental acts includes *dehumanization* and *attribution of blame*. Dehumanization divests people of human qualities or attributes bestial qualities to them. Once dehumanized, they are no longer viewed as persons with feelings, hopes, and concerns but as subhuman objects (Haritos-Fatouros, 1988; Keen, 1986; Kelman, 1973). It is difficult to mistreat humanized persons without risking personal distress and self-censure. By attribution of blame, people view themselves as faultless victims driven to injurious conduct by forcible provocation (Crick & Dodge, 1994; Darley, Klosson, & Zanna, 1978; Ferguson & Rule, 1983; Weiner, 1986). Punitive conduct thus becomes a justifiable defensive reaction to instigations.

A number of findings have shown that the above eight mechanisms can be traced back to a common latent variable that makes people more or less inclined to use mechanisms of moral disengagement (Bandura et al., 1996, 2001; Caprara, Bandura, Barbaranelli, & Vicino, 1996; Pelton, Gound, Forehand, & Brody, 2004). Thus, the proneness to moral disengagement is assessed by the processes through which it is presumed to operate. Appropriate assessment measures have already been developed in order to investigate how the full set of moral disengagement mechanisms operate in concert (Bandura, 2004; Bandura et al., 1996; Caprara et al., 1996).

A large body of research has demonstrated the disinhibitory power of moral disengagement in fostering aggressive behavior (Andrus, 1969; Bandura, 1990, 2006; Kelman & Hamilton, 1989; Rapoport & Alexander, 1982; Reich, 1990). In adolescence, different studies have shown strong associations between proneness to moral disengagement and various measures of aggression and violence, as well as other forms of antisocial conduct and bullying (Bandura, Caprara, & Zsolnai, 2000; Bandura et al., 1996, 2001; Caprara et al., 1996; Gini, 2006; Menesini et al., 2003). Longitudinal studies have found that high levels of moral disengagement predict levels of violence, theft, and other forms of antisocial conduct (Bandura et al., 2001; Elliott & Rhinehart, 1995; Pepler, Jiang, Craig, & Connolly, 2008). On the one hand, high moral disengagers tend to be more irritable, more prone to vengeful rumination, more inclined toward physical and verbal aggression, and more frequently involved in violent episodes. On the other hand, high moral disengagers are less troubled by anticipatory feelings of guilt due to injurious conduct, that is, the higher the moral disengagement, the weaker the felt guilt and the lower the need to undo any harm caused by aggressive conduct (Bandura et al., 1996). High moral disengagers not only experience low guilt over injurious conduct but they are also less prosocial and less able to resist peer pressure for transgressive activities (Bandura et al., 1996, 2001; Kwak & Bandura, 1997).

Moreover, boys are more likely than girls to become moral disengagers over the course of development, although gender differences do not exist in the earlier years (Bandura, 2004; Bandura et al., 1996). Some findings from cross-cultural studies indicate that some of the gender differences in aggression may reside in the differential proclivity of disengaging moral self-sanctions from injurious conduct (Bandura, 1999, 2004).

Aim of Present Study: A Developmental Perspective

Our general aim was to examine the predictive power of different moral disengagement developmental trends in adolescence that are conducive to different levels of later aggressive and violent behaviors. To this aim, we used longitudinal structural equation modeling (i.e., growth latent curve and latent class growth analysis), a procedure not typically used in research on moral development. In particular, we first investigated the development of moral disengagement at a population level using growth latent curve analysis (Duncan & Duncan, 1994, 1995; Duncan, Duncan, & Hops, 1994; McArdle, 1988; McArdle & Anderson, 1989; McArdle & Hamagami, 1991; Meredith & Tisak, 1990; Muthén & Khoo, 1998; Stoolmiller, 1994) among an Italian adolescent sample from 14 to 20 years old. We hypothesized a decline in moral disengagement due to social, emotional, and cognitive skills that, over the course of development, enhance youths' capacities to regulate their own affect, to commit themselves to moral principles that ban aggression and violence, to accord their behavior to social standards that decrease behaviors that offend others, and to interact effectively with others (Eisenberg, Fabes, & Spinrad, 2006; Hart & Carlo, 2005; Tangney et al., 2007).

A second goal was to identify subgroups of adolescents following distinct developmental patterns of moral disengagement using latent class growth analysis, a semiparametric group-based approach (Nagin, 1999, 2005) that should be seen as a natural extension of growth models (Hoeksma & Kelderman, 2006). We hypothesized four trajectories, in accordance with what has been previously found in the aggression and in the aggressogenic personality factors literature (Barker, Tremblay, Nagin, Vitaro, & Lacourse, 2006; Brame et al., 2001; Broidy et al., 2003; Caprara, Paciello, Gerbino, & Cugini, 2007; Côté, Tremblay, Nagin, Zoccolillo, & Vitaro, 2002; Nagin & Tremblay, 1999; Tremblay & Nagin, 2005). We hypothesized a "nondisengaged" group whose moral disengagement would start and remain low, a "chronic" group whose moral disengagement would start and remain high, and two trajectories that could either decrease ("desister") or increase ("escalator") over the course of adolescence. To validate the trajectory model, the groups were compared on aggressive and violent behavior and a proxy of guilt across time (i.e., need of reparation; see Bandura et al., 1996; Caprara, Barbaranelli, Pastorelli, Cermak, & Rosza, 2001; Caprara, Manzi, & Perugini, 1992). We hypothesized that the chronic group would show higher levels of aggressive and violent behavior and lower levels of guilt and that the nondisengaged group would show lower levels of aggressive and violent behavior and higher levels of guilt over time.

Our third goal was to explore how gender is associated with moral disengagement development. We hypothesized that boys are more likely to exhibit higher levels of moral disengagement than girls in the course of development (Bandura et al., 1996) and that the decline in moral disengagement would be stronger in girls than in boys due to the higher resistance of girls to resort to direct aggression (Archer, 2004). As a consequence, we hypothesized that girls would have a higher probability of being members of a "nondisengaged" group, whereas boys would demonstrate a higher probability of being members of a "chronic" or "escalator" group.

Our fourth goal was to explore how peer evaluation of aggression in early adolescence is associated with moral disengagement development. Because it is more probable that adolescents who have already shown more aggressive behaviors in early adolescence could have more opportunities to recourse to moral disengagement, we hypothesized that the more preadolescents are evaluated as aggressive by peers, the higher the level of moral disengagement, the less the likelihood of declining during adolescence, and thus, the higher the probability of being a member of a "chronic" or "escalator" group and the lower the probability of being a member of a "nondisengaged" group.

Our final goal was to explore the association of moral disengagement development with aggressive and violent conduct in late adolescence, controlling for earlier peer evaluations of aggression. We hypothesized that the earlier preadolescents are inclined to use moral disengagement, and the more persistent recourse to it, the higher the engagement in aggression and violence. As a consequence, adolescents with a chronic high or increasing trajectory would show higher levels of aggression and violence relative to nondisengaged adolescents.

Method

Sample and Data Collection

The participants were part of an ongoing longitudinal project that began in 1989. We implemented this project to investigate the main determinants and pathways of aggression and successful development and adjustment from late infancy to early adulthood. The longitudinal project followed a staggered, multiple cohort design, with two cohorts assessed at five different time points. The first cohort was age 12 in 1994, the second cohort was age 12 in 1996 (Time 1). These participants were retested at age 14 (Time 2), at age 16 (Time 3), at age 18 (Time 4), and finally, at age 20 (Time 5). Cohort effects were previously tested and were found to be insignificant for sociodemographic and major study variables.

Three hundred and sixty-six adolescents (177 boys and 189 girls) were included in the study. Participants attended sixth grade at Time 1 and eighth grade at Time 2. Most of them attended high school at Times 3 and 4, whereas at Time 5, the majority were college students (71%). At Time 1, the participating adolescents were drawn from two schools in a residential community located near Rome, Italy. The community was composed of families of skilled workers, farmers, professionals, local merchants, and their service staff. Occupational socioeconomic distribution matched the national profile (Istituto Italiano di Statistica, 2002). The composition of the families also matched national data with regards to type of families and number of children. Most subjects were from intact (94.1%) and, on average, one-child families. The participation rate was high during longitudinal data collection, with 84.4% of the original sample (Time 1) returning at the final assessment time (Time 5). In particular, there were three patterns of missing data. Complete data were available for 65% of males and 83% of females. Thirty percent of males and 16% of females had missing data in one of the five waves. Finally, 5% of males and 1% of females had missing data in two of the five waves.

Procedures

Young adolescents at Times 1 and 2 were administered a set of scales tapping different types of dimensions in their classrooms by two trained female experimenters and were asked to complete the scales individually. Before starting, the experimenters explained that their responses to the questionnaires would be absolutely confidential. When necessary, the experimenters offered the children clarifications on the dimensions being measured. At Times 3, 4, and 5, participants were contacted by phone and invited to participate in the study for which they received a small payment. A stringent consent procedure for the research was followed including at various stages, parents' consent and approval from school councils, while letting children decline their participation if they so chose.

Measures

Moral disengagement (Bandura et al., 1996; Caprara et al., 1996). From Time 2 to Time 5, moral disengagement was assessed by self-report. The scale assesses proneness to moral disengagement of different forms of detrimental conduct in diverse contexts and interpersonal relationships. The full set of 32 items, presented in the Appendix, tap the eight different mechanisms by which moral self-sanctions can be disengaged from transgressive conduct as hypothesized by Bandura (1990, 1999). For each of the 32 items, adolescents rated on a 5-point Likert-type scale, their degree of acceptance of moral exonerations for such conduct on an agreement continuum (from 1 = *agree not at all* to 5 = *completely agree).*

Physical and verbal aggression scale (Caprara & Pastorelli, 1993). At Time 1, physical and verbal aggression was assessed by peer nominations. Children made their nominations from the roster of classmates in their classroom. Because this is a highly stable community, children were thoroughly acquainted with each other. Children were provided with a booklet containing all classmates' names. The respondents selected the three classmates who most frequently exert each of three different forms of physical and verbal aggression (kick and hit or punch, insult other kids or call them names, and hurt other kids). The numbers of nominations that each child received were summed separately and divided by the number of classmates (as a consequence this variable ranged from 0 = *never nominated by the classmates* to 1 = *nominated by all classmates*). Cronbach's reliability coefficient was .94.

From Time 2 to Time 5 physical and verbal aggression was assessed by self-report. This scale measured behaviors aimed to physically and verbally hurting others. It includes three items for physical aggression (i.e., "I hurt others") and three items for verbal aggression (i.e., "I say bad things about others kids") using a 3-point scale (from 1 = *never* to 3 = *often*) at Time 2 and a 5-point scale (from 1 = *never* to 5 = *often*) from Time 3 to Time 5. Cronbach's reliability coefficients ranged from .74 to .82 for physical aggression and from .68 to .72 for verbal aggression.

Violence (Caprara, Mazzotti, & Prezza, 1990). From Time 3 to Time 5, violence was assessed by self-report. This scale included 11 items aimed to assess the extent to which adolescents engage in violent conduct. For each item, adolescents rated how often they have engaged in violent actions, such as fighting, vandalism, or weapon use on a 5-point scale (from 1 = *never* to 5 = *often*). "Have you participated in violent actions of 'gangs'?" and "Have you ever had the occasion to use violence when there are arguments?" are two sample items. Cronbach's reliability coefficients ranged from .87 to .90.

Need for reparation (Caprara et al., 1992). From Time 2 to Time 5, need for reparation was assessed by self-report. This scale included five items measuring proneness to experience feelings of remorse, embarrassment, disturbance, tension, and desire for justice that are linked to the need for reparation of guilt-eliciting acts (i.e., "I feel that I have to make up for the wrongs that I've done to others") using a 6-point scale (from 1 = *completely true for me* to 6 = *completely false for me*). The Cronbach coefficients for this scale ranged from .65 to .71.

Results

Preliminary Analysis: Longitudinal Factorial Analysis

As a preliminary analysis, we first examined the dimensionality of the moral disengagement scale by using the exploratory factor analysis (EFA) approach. Second, we examined the longitudinal factorial invariance of moral disengagement to test the degree to which the construct is measured similarly across waves (Hoyle & Smith, 1994; Motl, Dishman, Birnbaum, & Lytle, 2005).

In accordance with previous studies on moral disengagement (Bandura et al., 1996; Caprara et al., 1996), two alternative models for each wave were hypothesized: (a) an eight-factor solution, each factor representing a distinct mechanism of moral disengagement, and (b) a one-factor solution, consistent with previous studies.

To identify the best solution, we analyzed the eigenvalues (Cattell & Vogelmann, 1977) and considered the standardized root mean square residual (SRMR; Jöreskog & Sörbom, 1984) and the root mean square error of approximation (RMSEA; Steiger & Lind, 1980) as indices of goodness of fit. Because several items presented a deviation from a normal distribution, EFA was performed using maximum likelihood parameter estimates with standard errors and the chi-square test statistic, which are robust to nonnormality (MLMV). Factors were rotated using the Promax procedure.

For Time 2, the first 10 eigenvalues of the correlation matrix were: 7.19, 2.21, 1.54, 1.36, 1.26, 1.12, 1.02, 0.98, and 0.94. The eight-factor solution did not converge. The analysis of eigenvalues suggested a one-factor solution. This model fit the data ($\chi^2 = 1192.19$, $df = 464$, $p < .001$, RMSEA = .06, SRMR = .07), reproducing with a good approximation the intercorrelation among the 32 items of the scale. Cronbach's reliability coefficient for this wave was .87.

For Time 3, the first 10 eigenvalues of the correlation matrix were: 10.67, 2.17, 1.68, 1.40, 1.26, 1.07, 0.99, 0.91, 0.88, and 0.82. The eight-factor solution did not converge. The analysis of eigenvalues suggested a one-factor solution. This model fit the data ($\chi^2 = 1291.78$, $df = 464$, $p < .001$, RMSEA = .07, SRMR = .07), reproducing with a good approximation the intercorrelation among the 32 items of the scale. Cronbach's reliability coefficient for this wave was .93.

For Time 4, the first 10 eigenvalues of the correlation matrix were: 9.99, 2.20, 1.61, 1.42, 1.31, 1.24, 1.04, 0.99, 0.96, and 0.90. Although the RMSEA and the SRMR of the eight-factor model resulted in a good approximation to the data (RMSEA = .03, SRMR = .03), the pattern of rotated factor loadings was not interpretable or simple (Thurstone, 1947). The analysis of eigenvalues suggested a one-factor solution. This model fit the data ($\chi^2 = 1322.68$, $df = 464$, $p < .001$, RMSEA = .07, SRMR = .07), reproducing with a good approximation the intercorrelation among the 32 items of the scale. Cronbach's reliability coefficient for this wave was .92.

For Time 5, the first 10 eigenvalues of the correlation matrix were: 9.15, 2.19, 1.95, 1.51, 1.39, 1.19, 1.08, 1.01, 0.90, and 0.87. Although the RMSEA and the SRMR of the eight-factor model resulted in a good approximation to the data (RMSEA = .03, SRMR = .03), the pattern of rotated factor loadings was not interpretable or simple (Thurstone, 1947). The analysis of eigenvalues suggested a one-factor solution. This model fit the data ($\chi^2 = 1293.32$, $df = 464$, $p < .001$, RMSEA = .08, SRMR = .09). Cronbach's reliability coefficient for this wave was .91.

The above findings substantially corroborated previous results (Bandura et al., 1996; Caprara et al., 1996) and attest to a main latent dimension, which includes all the items measuring the eight moral disengagement mechanisms. As a next step, we analyzed the longitudinal factor structure of moral disengagement across the four waves of data (14–20 years) testing a series of nested models, from less to more demanding ones. As Meredith (1993) and Widaman and Reise (1997) have argued, there are different levels of invariance. *Configural invariance* (Thurstone, 1947) hypothesizes the equality of the overall structure (i.e., same factor and same patterns of fixed and freed parameters) across groups or time. *Metric* or *pattern invariance* (Thurstone, 1947) hypothesizes the equality of the factor loadings across time or groups. *Strong factorial invariance* (Meredith, 1993) hypothesizes the equality of the intercept of the measured variables across time or groups. *Strict factorial invariance* (Meredith, 1993) hypothesizes the equality of the uniqueness of the measured variables across time or groups.

Given the small size of the longitudinal sample and the number of indicators of moral disengagement, the longitudinal invariance was analyzed via item parceling (Anderson & Gerbing, 1988; Hoyle 1995; Martinez, Black, & Starr, 2002). Parceling typically exhibits a distribution that more closely approaches a normal distribution than using the original items as individual indicators (West, Finch, & Curran, 1995). Thus, with the aim of reducing the variance–covariance matrix to be analyzed, we built four parcels for each wave. We used item-scale correlation to select items. In particular, items were assigned to parcels so that equivalent parcels in terms of homogeneity were created, and the monodimensionality of each parcel was guaranteed as suggested by Bandalos and Finney (2001).

In all models, autocorrelations were estimated among all pairs of uniqueness because the same parcels were used across the four time points (Motl et al., 2005; Pitts, West, & Tein, 1996). To compare the fit of the nested models in the longitudinal invariance sequence, we used the chi-square difference test (Bollen, 1989; Fix, Hodges, & Lehmann, 1959) and the difference in comparative fit index (CFI; Bentler, 1989); as noted by Cheung and Rensvold (2002), a difference larger than .01 in CFI would indicate a meaningful change in model fit.

The fit indices for the models tested in the invariance analyses are presented in Table 1. As can be noted, configural and metric invariance were supported by all fit indices. Although full strong invariance was not supported, support for partial strong invariance came from differences in the CFI. Finally, support for a strict factorial invariance only came from differences in the CFI. Because the majority of loadings were invariant, results confirmed that the constructs were measured as comparable over time (Raykov, 2004; Raykov & Marcoulides, 2006).

Descriptive Statistics

Observed means of moral disengagement from Time 2 to Time 5, peer evaluation of physical and verbal aggression measured at Time 1, and the outcomes (violence and physical and verbal aggression) measured at Time 5, separately for males and females, are reported in Table 2.

As shown, there were missing data in all the variables. As noted by Hanson, Tobler, and Graham (1990), in longitudinal research, it is common to have subject attrition, which produces missing values in the data set. In the presence of missing data, estimation of parameters must be adjusted accordingly. Among the different methods used for taking into account missing data, we used the maximum likelihood estimation of parameters, a method widely accepted as appropriate for handling missing data (Muthén & Shedden, 1999; Schafer & Graham, 2002) under the assumption that the data are missing at random (Arbuckle, 1996; Little, 1995).

Before proceeding with the analysis, the normality of the variables was ascertained. Due to the non-normality of some measures (the skewness and kurtosis varied from 2.23 for physical aggression to 10.79 for violence), we computed the inverse of violence and physical aggression and the root square of peer evaluation of physical and verbal aggression to normalize these variables as suggested by Tabachnik and Fidell (1989). The skewness and kurtosis of the computed outcomes varied from −.572 for physical aggression to −1.29 for violence.

Correlations between moral disengagement across the four time points, peer evaluation of physical and verbal aggression, and the three outcomes—physical and verbal aggression and violence—for boys and girls are provided in Table 3. As shown, they attested to a high-medium stability of moral disengagement

Table 1
Fit Indices for the Model of the Invariance Routine

Model	χ^2	*df*	*p*	CFI	RMSEA	SRMR					
Baseline											
14 years	.806	2	.67	1.00	.00 (.00 – .08), $p = .84$	.005					
16 years	.877	2	.65	1.00	.00 (.00 – .08), $p = .82$	.003					
18 years	3.696	2	.16	1.00	.05 (.00 – .12), $p = .40$	.007					
20 years	1.591	2	.45	1.00	.00 (.00 – .10), $p = .67$	.006					
								$\Delta\chi^2$	*df*	*p*	ΔCFI
Longitudinal invariance											
Configural	80.993	74	.27	1.00	.02 (.00 – .03), $p = 1.00$	.024					
Metric	95.223	83	.17	1.00	.02 (.00 – .04), $p = 1.00$	.040	Metric versus configural	14.23	9	.11	.00
Full strong	508.925	95	.00	0.91	.11 (.10 – .12), .00	.254	Full strong versus metric	412,80	12	.00	.09
Partial strong	151.322	89	.00	0.99	.04 (.03 – .05), $p = .80$	.058	Partial strong versus full strong	56.10	6	.00	.01
Strict	174.952	101	.00	0.98	.04 (.03 – .06), $p = .78$	.062	Strict versus partial strong	23,63	12	02	.01

Note. CFI = comparative fit index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual.

Table 2

Observed Descriptive Statistics of Moral Disengagement, Predictor and Outcomes Separately for Gender

	Males			Females		
	N	*M*	*SD*	*N*	*M*	*SD*
MD14	174	2.68	0.45	187	2.50	0.46
MD16	168	2.34	0.63	178	2.02	0.52
MD18	172	2.23	0.55	188	1.90	0.50
MD20	137	2.11	0.50	172	1.81	0.41
AGG_PE	166	0.15	0.19	188	0.05	0.07
PHY	137	1.53	0.71	172	1.26	0.50
VERB	137	1.97	0.75	172	1.61	0.58
VIO	136	1.40	0.52	170	1.11	0.21

Note. MD14 = moral disengagement at age 14; MD16 = moral disengagement at age 16; MD18 = moral disengagement at age 18; MD20 = moral disengagement at age 20; AGG_PE = physical and verbal aggression peer evaluation at age 12; PHY = physical aggression at age 20; VERB = verbal aggression at age 20; VIO = violence at age 20.

over time with lower correlations the longer the distance of time. Finally, peer evaluation of physical and verbal aggression and the outcomes were correlated with the moral disengagement variable, especially in boys.

Growth Models

The analysis of moral disengagement development was carried out within a latent variable framework. We specified a multigroup growth curve model that simultaneously estimated the same pattern of relationships among variables for males and females. To examine gender differences in the estimated parameters, we constrained all parameters to be equal across groups and used the chi-square difference test to compare nested models. Modification indices were used to assess the tenability of the equality constraint imposed across gender.

Table 3

Correlations Between Moral Disengagement From Ages 14 to 20, Physical and Verbal Aggression at Age 12, and Aggressive and Violent Conduct at Age 20, Separately for Gender

	1	2	3	4	5	6	7	8
1. AGG_PE		*.09*	*.11*	.27	.21	*.12*	.16	.17
2. MD14	.21		.51	.40	.43	*.11*	*.01*	*.05*
3. MD16	.29	.49		.66	.52	.17	.15	.21
4. MD18	.21	.44	.58		.64	.22	.21	.22
5. MD20	.18	.39	.37	.65		.41	.32	.42
6. VERB	*.15*	.39	.25	.37	.50		.33	.41
7. PHY	.16	.31	.32	.42	.53	.67		.57
8. VIO	.25	.21	.32	.37	.40	.45	.53	

Note. Values below the diagonal are for males. Values above the diagonal are for females. All correlation coefficients were significant at least at $p < .05$, except for those that are in italics, which are not significant. Agg_PE = physical and verbal aggression peer evaluation at age 12; MD14 = moral disengagement at age 14; MD16 = moral disengagement at age 16; MD18 = moral disengagement at age 18; MD20 = moral disengagement at age 20; PHY = physical aggression at age 20; VERB = verbal aggression at age 20; VIO = violence at age 20.

We defined two latent growth factors from multiple indicators, that is, the four repeated measures of moral disengagement. The first factor was labeled intercept and represents the initial starting point of moral disengagement. The second factor represents the slope or the shape of the trajectory over time. The following equation shows the mathematical representation of the growth model:

$$y_t = \eta_0 + \eta_1 x_t + \varepsilon_t; \quad \text{where } t = \text{Times } 2, 3, 4, 5,$$

where y_t = observed score at time t, η_0 = unobserved score for the intercept factor, η_1 = unobserved score for the growth rate factor, and x_t = factor loading relating y_t to latent growth variables.

Because the factor loadings of the slope give the shape of the growth, a series of models were tested and compared with each other. In this way, we could determine the parameterization that best fit the data. In all the models tested, we defined the intercept as moral disengagement at age 14 by fixing the factor loading relating this variable to the slope at 0.

Following McArdle and Anderson (1989), the first model tested was a no-growth model (only intercept), $\chi^2(14, N = 177, 188) = 419.497$, $p < .001$, CFI = .08, RMSEA = .398 (.366, .482), SRMR = .779. This model assumed that the level of moral disengagement was stable over time except for a random error component at each time of evaluation. The second was a linear model representing a constant change over time, $\chi^2(8, N = 177, 188) = 70.668$, $p < .001$, CFI = .86, RMSEA = .207 (.164, .253), SRMR = .077. In this model, we fixed the factor loadings on the slope at 0, 1, 2, and 3. The third model examined a nonlinear growth where the form of the change over time was not specified a priori, $\chi^2(6, N = 177, 188) = 4.086$, $p = .67$, CFI = 1.00, RMSEA = .000 (.000, .077), SRMR = .040. Finally, the fourth model was the nonlinear model without the intercept, $\chi^2(12, N = 177, 188) = 88.738$, $p < .001$, CFI = .83, RMSEA = .187 (.152, .225), SRMR = .230. In particular, we fixed the first factor loadings on the slope and estimated the others. In all the growth models, autocorrelations between the same measures over time were estimated and for model identification, two uniquenesses and two covariances between uniquenesses were constrained to be equal among males and females. Because these models are nested, we performed a chi-square differences test to compare the models (Bollen, 1989; Fix et al., 1959). This test revealed that the nonlinear model provided the best fit to the data compared to the no-growth model, $\Delta\chi^2(8) = 415.41$, $p < .001$; the linear growth model, $\Delta\chi^2(2) = 66.58$, $p < .001$; and the nonlinear without intercept model, $\Delta\chi^2(6) = 84.65$, $p < .001$. Then, we tested if there were any differences among males and females on the free factor loadings, constraining these parameters to be equal among the two groups and used the chi-square difference test to compare the constrained model with the nonconstrained one. Results showed that the free factor loadings on the slope ($x_4 = 1.282$, $x_5 = 1.513$) were the same for males and females, $\Delta\chi^2(2) = 1.57$, $p = .46$, and were significantly smaller than the linear loadings ($x_4 < 2$ with $p < .001$, $x_5 < 3$ with $p < .001$), indicating that the decrease in moral disengagement from ages 14 to 16 was larger than both the decrease from ages 16 to 18 and the decrease from ages 18 to 20.

Table 4
Growth Curve Parameters

	Males		Females	
Growth parameters	Effect	*t* value	Effect	*t* value
Mean				
Intercept[a]	2.68	78.25	2.51	75.51
Slope[a]	−0.36	−11.67	−0.46	−13.61
Variances				
Intercept	0.19	3.66	0.19	3.66
Slope	0.08	2.61	0.08	2.61
Correlation				
Intercept ↔ slope	−.061	−2.06	−0.61	−2.06

Note. The *t* values greater than 1.96 (1.65 for variances) in magnitude indicate a parameter estimate that is significantly different from zero (for $p < .05$). Parameters estimated for correlation (↔) are presented in standardized form. All other parameter estimates are presented in unstandardized form.
[a] Parameter is different for males and females.

The results of the final model, $\chi^2(9, N = 177) = 5.096$, 188, $p = .83$, CFI = 1.00, RMSEA = .000, CI = .000, .050, SRMR = .065, summarized in Table 4 and Figure 1, suggest that there were gender differences in the initial level of moral disengagement; that is, males exhibited higher levels of moral disengagement than females at age 14—we compared the model with the mean of the intercept constrained to be equal among males and females with the nonconstrained model, $\Delta\chi^2(1) = 13.03$, $p < .001$. Furthermore, the mean of the slope indicated a significant decrease from ages 14 to 16. In particular, females exhibited a greater decline than males—we compared the model with the mean of the slope constrained to be equal among males and females with the nonconstrained model, $\Delta\chi^2(1) = 7.64$, $p < .001$. The variances of the growth factors were also estimated and they suggested that there was a significant variation in individual differences in the initial status and in the growth rate and that these were the same for both males and females—we compared the model with the variance of the intercept and the variance of slope constrained to be equal among males and females with the nonconstrained model, $\Delta\chi^2(2) = .08$, $p = .96$. Finally, results demonstrate that the covariance between the initial level and the slope was significantly negative and equal for males and females, that is, the higher the level of moral disengagement at age 14, the less the change of moral disengagement during adolescence—we compared the model with the covariance among intercept and slope constrained to be equal among males and females with the nonconstrained model, $\Delta\chi^2(1) = .59$, $p = .44$.

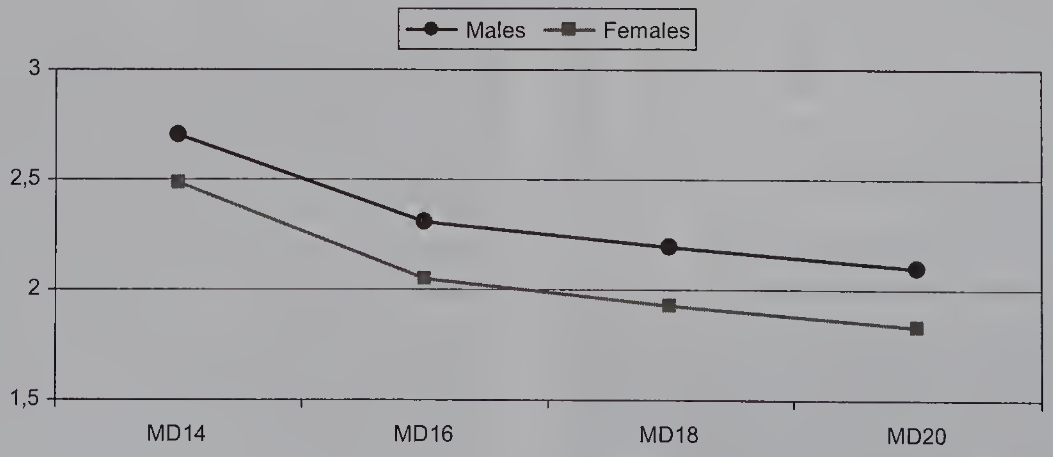


Figure 1. Moral disengagement development.
Note. MD14 = moral disengagement at age 14; MD16 = moral disengagement at age 16; MD18 = moral disengagement at age 18; MD20 = moral disengagement at age 20.

After ascertaining the best fitting growth curve, we added the covariate peer evaluation of physical and verbal aggression at age 12 and the three outcomes to examine the influence exerted by the covariate on moral disengagement at age 14, on its development during adolescence, and on physical and verbal aggression and violence at age 20 separately for males and females. Furthermore, we examined the influence exerted by moral disengagement at age 14 and its development during adolescence on physical and verbal aggression and violence at age 20. The fit indices of the model suggested a good approximation of the model to the data. In particular, $\chi^2(34, N = 177, 188) = 28.077$, $p = .75$, CFI = 1, RMSEA = .000 (.000 – .039), SRMR = .082.

The results of the regression part of this model, summarized in Table 5, indicate that peer evaluation of physical and verbal aggression at age 12 significantly predicted a higher level of moral disengagement at age 14 for both males and females. Peer evaluation of physical and verbal aggression at age 12 affected physical and verbal aggression at Time 5 only indirectly through moral disengagement at age 14 (intercept; for physical aggression: male, $\beta = .17$, $z = 3.415$; female, $\beta = .11$, $z = 2.647$; for verbal aggression: male, $\beta = .18$, $z = 3.472$; female, $\beta = .12$, $z = 2.783$). Moreover, peer evaluation of physical and verbal aggression at age 12 directly affected violence at Time 5 for males only and indirectly through moral disengagement at age 14 (intercept) both for males and females (male: $\beta = .11$, $z = 3.421$; female: $\beta = .14$, $z = 2.954$). Furthermore, moral disengagement measured at age 14 significantly predicted physical and verbal aggression and violent behavior 6 years later for both males and females. In particular, the higher the level of moral disengagement at age 14, the higher the levels of physical and verbal aggression and violence problems at age 20. In addition, results suggested that the more moral disengagement decreased from age 14 to 16, the lower the expected levels of physical and verbal aggression and violence at age 20 for both males and females. The model accounted for 33% and 24% of the variance of verbal aggression, 34% and 25% of the variance of physical aggression, and 28% and 34% of the variance of violence for males and females, respectively.

Table 5
The Effect of the Growth Curve of Moral Disengagement on Aggressive Behavior

	Males		Females	
	Effect	*t* value	Effect	*t* value
Correlation				
VIO ↔ PHY	.18	4.17	.28	4.17
VIO ↔ VERB	.12	3.20	.21	3.20
VERB ↔ PHY[a]	.28	3.97	.18	2.52
Regression				
Agg_Pe → Intercept[a]	.26	3.98	.25	3.33
Agg_Pe → Slope	.00	1.69	.14	1.68
Agg_Pe → PHY	.00	0.35	.02	0.35
Agg_Pe → VERB	.00	0.01	.00	0.01
Agg_Pe → VIO[a]	.21	2.92	.00	0.03
Intercept → PHY[a]	.66	6.58	.42	4.44
Slope → PHY	.54	5.42	.60	5.42
Intercept → VERB[a]	.67	6.90	.47	5.15
Slope → VERB	.49	5.60	.59	5.60
Intercept → VIO	.39	6.54	.54	6.54
Slope → VIO	.49	5.48	.69	5.48
Agg_Pe → VIO	.16	2.20	.00	0.07

Note. The values greater than 1.96 (1.65 for variances) in magnitude indicate a parameter estimate that is significantly different from zero (for $p < .05$). Parameters estimated for correlation (↔) and regression (predictor → outcome) are presented in standardized form. PHY = physical aggression at age 20; VERB = verbal aggression at age 20; VIO = violence at age 20; Agg_Pe = physical and verbal aggression peer evaluation at age 12.
[a]Parameter is different for males and females.

Trajectory Model

A semiparametric group-based approach (Nagin, 1999, 2005) was then used to identify developmental trajectories. The Bayesian information criterion (BIC; Schwarz, 1978) was used to determine the optimal number of groups. This method allows: (a) the identification of subgroups of individuals following distinct developmental trajectories for a given dimension, (b) the estimation of the proportion of individuals following each trajectory, and (c) the estimation of the stability of each subgroup over time (i.e., the shape of each trajectory over time). Censored normal models were estimated. A detailed description of the statistical rationale underlying the trajectory estimation procedure is given elsewhere (e.g., Jones & Nagin, 2007; Jones, Nagin, & Roeder, 2001; Nagin, 1999, 2005).

The analysis proceeded in four stages. First, the optimal number of groups was identified and models with various specifications for stable, linear, quadratic, or cubic shapes of the moral disengagement trajectory groups were estimated. In the second stage, after ascertaining the best trajectory model, we tested whether and how much gender and peer evaluation of physical and verbal aggression at age 12 affected the probability of group membership. In the third stage, we validated the trajectory model examining the profiles of each group on four correlates across time. In particular, we considered physical aggression, verbal aggression, violence, and need for reparation. Finally, the linkage between trajectories of moral disengagement and aggressive and violent outcomes at age 20 was tested (Nagin, 1999).

First stage: Identification of the developmental model. A four-group unconditional model showed the best fit to the data (BIC = −867.58). The four-group model was respecified considering gender and peer evaluation of physical and verbal aggression as predictors of the trajectory group membership probability. The four-group conditional model was confirmed as the best model (BIC = −832.89). Figure 2 presents the size and the shapes of the trajectory groups.

The developmental trajectory model suggested the following groups:

- Nondisengaged group: composed of 37.9% of the adolescents characterized by a quadratic trend with initially low levels of moral disengagement followed by a significant decline;
- Normative group: composed of 44.5% of the adolescents characterized by a cubic trend with initially moderate levels of moral disengagement followed by a significant decline;
- Later desister group: composed of 6.9% of the adolescents characterized by a cubic trend with initially high-medium levels of moral disengagement followed by a significant increase from ages 14 to 16 and an even more significant decline from ages 16 to 20;
- Chronic group: composed of 10.7% of the adolescents with constant medium-high levels of moral disengagement.

Table 6 presents the correlations between the actual grouping variables (dichotomous) and the posterior group membership probabilities.

With regard to the gender distribution of the group trajectories, there were a higher proportion of boys in the chronic group and a higher proportion of girls in the nondisengaged group (see Figure 2).

Second stage: Linking group membership to gender and aggressive predictor. Table 7 presents the results of the multinomial logit regression used to examine whether and how much gender and peer evaluation of physical and verbal aggression affected the probability of group; membership. The first panel of Table 4 shows coefficient estimates and *t* statistics. The normative group served as the contrast group. Relative to the normative group, gender significantly increased the probability of membership in the nondisengaged group; that is, females had a higher probability of being in this group than did males. Furthermore, higher levels of peer evaluation on physical and verbal aggression significantly increased the probability of membership in the chronic and later desister groups (α = .05, one-tailed test).

The second panel of Table 7 shows the predicted probabilities of group membership based on the coefficient estimates. Both males and females who were never nominated by their classmates had a high probability of belonging to the normative or

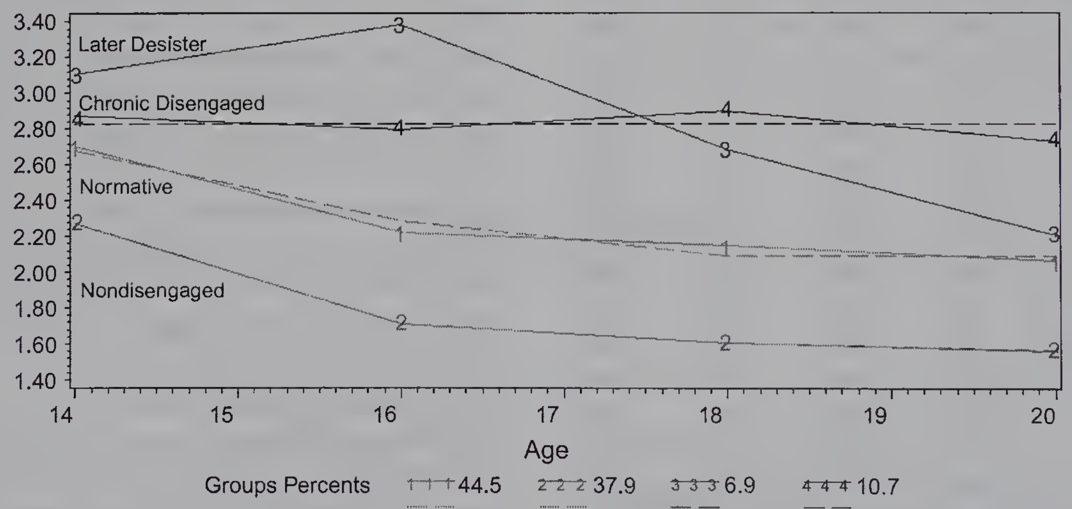


Figure 2. Moral disengagement trajectories.

Note. The continuous lines represent trajectories of actual moral disengagement calculated as mean scores for adolescents in groups identified by Proc traj procedure. The broken lines represent predicted moral disengagement calculated with model's coefficient estimates. Gender distribution in each trajectory group: nondisengaged group: 23.1% male and 76.9% female (of the total sample, 18.7% male and 54.8% female); normative group: 54.7% male and 45.3% female (of the total sample, 52.4 % male and 38.3% female); chronic disengaged group: 89.7% male and 10.3% female (of the total sample, 21.1% male and 2.1% female); later desister group: 59.1% male and 40.9% female (of the total sample, 7.8% male and 4.8% female).

Table 6

Correlation Between the Actual Grouping Variables and the Posterior Group Membership Probabilities

	Normative PRB	Nondisengaged PRB	Later desister PRB	Chronic PRB
Normative	.94	−.62	−.24	−.28
Nondisengaged	−.65	.95	−.26	−.32
Later desister	−.25	−.23	.95	*−.01*
Chronic disengaged	−.30	−.32	*.05*	.95

Note. All correlation coefficients were significant at least at $p < .05$, except for those in italics, which are not significant. PRB = posterior probabilities.

nondisengaged groups (.86 and .96, respectively), but while females had a higher probability of being a member of the nondisengaged group (.61), males had a higher probability of being a member of the normative group (.55).

Considering a high-risk scenario in which adolescents were nominated by all classmates, the predicted probability for males of being a member of the chronic group was .38, whereas it was .10 in the previous scenario. In a similar way, the predicted probability for females of being a member of the chronic group was .14, whereas it was .02 in the previous scenario. Furthermore, both males and females had a high-predicted probability to belong to the later desister or normative group (.58 and .70, respectively). Finally, the predicted probability of males being a member of the nondisengaged group was .04 instead of .31 in the previous scenario. In a similar way, the predicted probability for females to be a member of the nondisengaged group was .15, whereas it was .61 in the previous scenario.

Third stage: Trajectories profiles on aggression, violence and need for reparation. Figure 3 shows the profiles in physical aggression, verbal aggression, violent conduct, and need for reparation for each group across time.

As shown, the four trajectory groups significantly differed on physical aggression (Duncan post hoc test). At ages 14, $F(3, 229) = 40.02$, $p < .000$, $\eta^2 = .21$, and 16, $F(3, 317) = 42.47$, $p < .001$, $\eta^2 = .29$, all four groups were different from each other. In particular, the later desister group showed the highest level of physical aggression, whereas the nondisengaged group showed the lowest level. At age 18, $F(3, 330) = 44.94$, $p < .001$, $\eta^2 = .29$, the nondisengaged group showed the lowest level of physical aggression, whereas the later desister and chronic groups did not differ from each other and showed the highest level. Finally, at age 20, $F(3, 299) = 22.02$, $p < .001$, $\eta^2 = .18$, the chronic group showed the highest level of physical aggression.

With regard to verbal aggression, the four trajectory groups were significantly different (Duncan post hoc test). In particular, at age 14, $F(3, 330) = 20.01$, $p < .000$, $\eta^2 = .15$, the later desister and chronic groups did not differ from each other and showed the highest level of verbal aggression. At age 16, $F(3, 317) = 37.38$, $p < .001$, $\eta^2 = .26$, the later desister and chronic groups did not differ from each other and showed the highest level of verbal aggression and the nondisengaged showed the lowest. At age 18, $F(3, 330) = 42.53$,

Table 7

The Impact of Gender and Physical and Verbal Aggression Peer Evaluated on Group Membership Probabilities

	Groups			
	Later desister disengaged	Chronic disengaged	Normative	Nondisengaged
	Multinomial logit coefficients (with ± statistics given in parenthesis)			
Constant	−2.67 (−3.87)	−2.84 (−3.45)	—	0.56 (1.86)
Gender	−0.05 (0.09)	−1.15 (−1.37)	—	−1.12 (−3.38)
Agg_Pe	2.78 (2.11)	2.01 (1.89)	—	−1.38 (−1.46)
	Predicted membership probabilities based on multinomial model coefficient estimates			
Males and Agg_Pe = 0	.04	.10	.55	.31
Females and Agg_Pe = 0	.02	.02	.35	.61
Males and Agg_Pe = 1	.30	.38	.28	.04
Females and Agg_Pe = 1	.37	.14	.33	.15

Note. Agg_Pe = physical and verbal aggression peer evaluation at age 12. Agg_Pe = 0: *never nominated by classmates*; Agg_Pe = 1: *nominated by all classmates.*

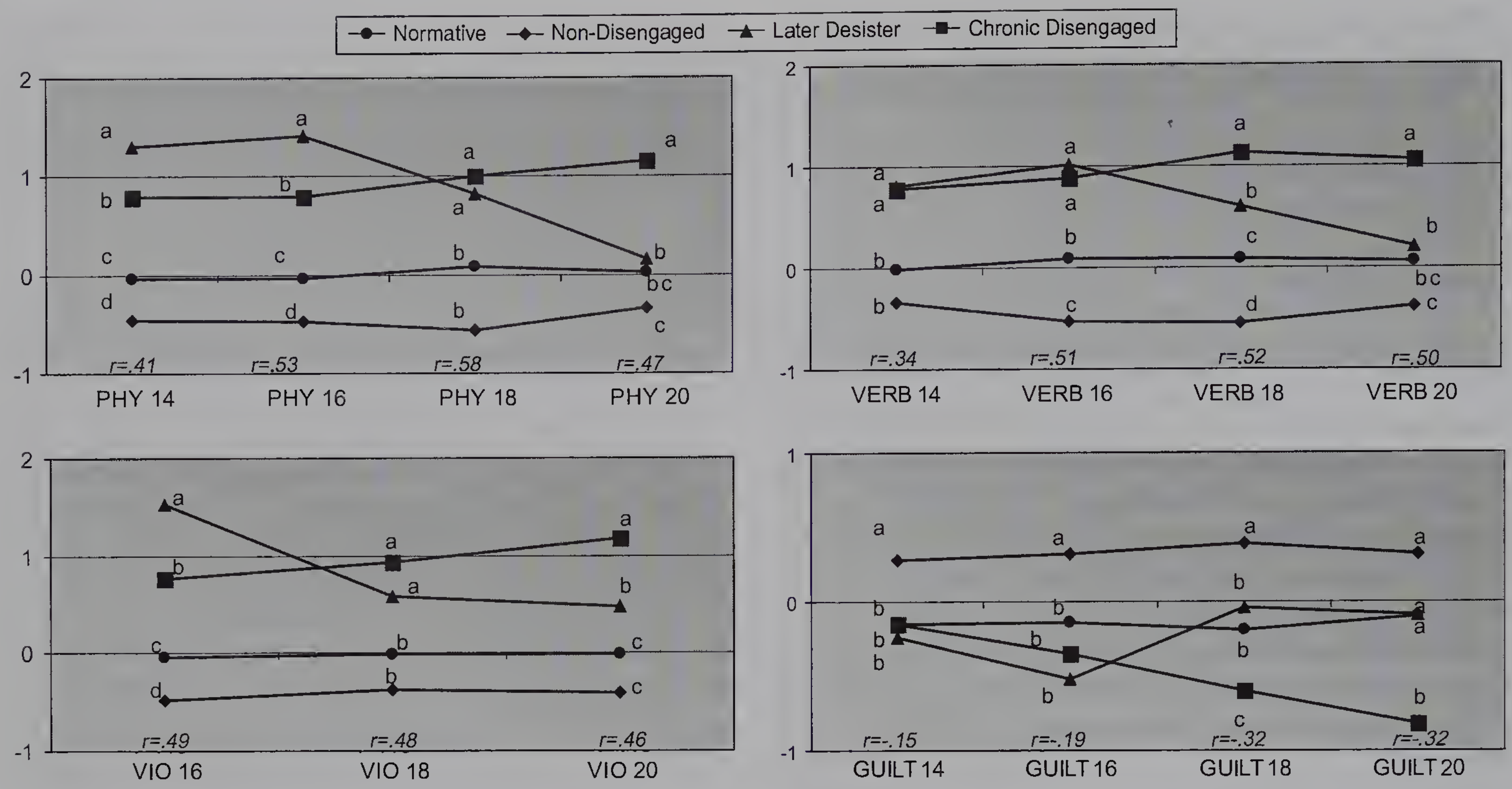


Figure 3. Trajectories profiles on aggressive behavior, violence, and need for reparation.
Note. PHY14 = physical aggression at age 14; PHY16 = physical aggression at age 16; PHY18 = physical aggression at age 18; PHY20 = physical aggression at age 20; VERB14 = verbal aggression at age 14; VERB16 = verbal aggression at age 16; VERB18 = verbal aggression at age 18; VERB20 = verbal aggression at age 20; VIO16 = violence at age 16; VIO18 = violence at age 18; VIO20 = violence at age 20; GUILT14 = need for reparation at age 14; GUILT16 = need for reparation at age 16; GUILT18 = need for reparation at age 18; GUILT20 = need for reparation at age 20. In each figure correlations between moral disengagement and each variable are reported for each time point.

$p < .001$, $\eta^2 = .28$, all four groups were different from each other. In particular, the chronic group showed the highest level of verbal aggression and the nondisengaged the lowest. Finally, at age 20, $F(3, 299) = 21.53$, $p < .001$, $\eta^2 = .18$, the chronic group showed the highest level of verbal aggression, whereas the normative and later desister groups did not differ from each other.

With regards to violence, the four trajectory groups were significantly different (Duncan post hoc test). In particular, at age 16, $F(3, 317) = 48.02$, $p < .000$, $\eta^2 = .31$, all four groups were different from each other. In particular, the later desister group showed the highest level of violence. At age 18, $F(3, 330) = 24.42$, $p < .001$, $\eta^2 = .18$, the chronic and later desister groups did not differ from each other and showed the highest level of violence. Finally, at age 20, $F(3, 286) = 27.33$, $p < .001$, $\eta^2 = .22$, the chronic group showed the highest level of violence, whereas the later desister group showed higher levels of violence than the normative and nondisengaged groups.

Finally, the four trajectory groups significantly differed on need for reparation (Duncan post hoc test). In particular, at ages 14, $F(3, 246) = 5.44$, $p < .000$, $\eta^2 = .05$, and 16, $F(3, 327) = 8.92$, $p < .001$, $\eta^2 = .08$, the nondisengaged group showed the highest level of need for reparation. At age 18, $F(3, 342) = 15.02$, $p < .001$, $\eta^2 = .12$, the nondisengaged group showed the highest level and the chronic showed the lowest. Finally, at age 20, $F(3, 299) = 12.60$, $p < .001$, $\eta^2 = .11$, the chronic group showed the lowest level of need for reparation, whereas the other three groups did not differ.

In each figure, the correlations between moral disengagement and each variable were reported. As shown, moral disengagement was positive and highly correlated with physical and verbal aggression and violence and negatively correlated with need of reparation in each time point.

Fourth stage: The impact of moral disengagement trajectories on aggressive and violent conduct. Table 8 presents the correlations between the posterior probabilities of group membership for each individual and aggressive and violent conduct at age 20.

As shown, aggressive and violent outcomes were positively and significantly correlated with the posterior probability of being in the chronic or later desister groups. On the contrary, the outcomes were negatively and significantly correlated with the

Table 8

Correlations Between the Posterior Probabilities of Group Membership and Physical and Verbal Aggression and Violence at Age 20

Posterior probabilities	Verbal aggression	Physical aggression	Violence
Normative	*.07*	*.04*	*.10*
Nondisengaged	−.35	−.33	−.42
Later desister	.10	.17	.20
Chronic	.39	.36	.39

Note. All correlation coefficients were significant at least at $p < .05$, except the one that is underlined, which is significant at $p < .10$, and those that are italics, which are not significant.

posterior probability of being in the nondisengaged group. Moreover, correlations were not significant between outcomes and the posterior probability of being in the normative group.

To test the linkage between trajectories of moral disengagement and the three outcomes (i.e., physical aggression, verbal aggression, and violent behaviors at age 20), three hierarchical regression analyses were conducted. In these models, gender, peer evaluation on physical and verbal aggression at age 12, and the posterior probabilities of group membership were considered independent variables. In entering the posterior probabilities, we excluded the probability of group membership for one of the groups because the posterior probabilities of group membership in each group added to 1 for a given individual (e.g., the independent variables would be perfectly correlated). We decided to exclude the probability of being in the normative group.

As a preliminary step, we evaluated whether these analyses needed to be conducted separately for males and females. In this regard, the interaction of Gender × Peer Evaluation on Physical and Verbal Aggression and the interaction of Gender × The Different Group Membership Probabilities were included as independent variables in the hierarchical regression model. In particular, gender was entered at Step 1, peer evaluation of physical and verbal aggression was entered at Step 2, the interaction of Gender × Peer Evaluation of Physical and Verbal Aggression was entered at Step 3, the different group membership probabilities were entered at Step 4, and the interaction of Gender × The Different Group Membership Probabilities were entered at Step 5. The analysis revealed no statistically significant interactions; therefore, our subsequent hierarchical regression analyses were performed with gender as a factor rather than separately for the two gender groups.

The final hierarchical regression models are presented in Table 9. The results of the three hierarchical models showed that gender significantly predicted the three outcomes, especially violence (17% of explained variance). Moreover, peer nomination of physical and verbal aggression at age 12 significantly predicted higher levels of physical and verbal aggression and violent conduct at age 20 (the explained variance ranged from 2% to 5%).

In addition to the contribution of gender and earlier peer evaluations of aggressive behavior, the posterior probability of group membership significantly predicted the three outcomes (the explained variance ranged from 10% to 13%). In particular, the probability of being in the chronic group was positively and significantly linked with verbal aggression, physical aggression, and violence, whereas the probability of being a member of the nondisengaged group was negatively and significantly linked with aggression and violence.

Table 9

Results of Hierarchical Regression of Trajectories on Physical and Verbal Aggression and Violence at Age 20

		Verbal aggression				Physical aggression				Violence			
Step		β_1	β_2	β_3	R^2 change	β_1	β_2	β_3	R^2 change	β_1	β_2	β_3	R^2 change
1	Gender	.16***	.19**	*ns*	.07***	.23***	.15**	*ns*	.05***	.41***	.30***	.18***	.17***
2	Aggr_Pe		.16**	*ns*	.02***		.20***	.11†	.04***		.25***	.16***	.05***
3	Nondisengaged			−.20***	.13***			−.18**	.11***			−.22***	.10***
	Chronic disengaged			.28***				.25***				.20**	
	Later desister			*ns*				*ns*				*ns*	
	R^2				*.21*				*.19*				*.32*

Note. β_1 = beta coefficients before entering trajectories probabilities and levels of outcomes at Time 1; β_2 = coefficients before entering trajectories probabilities; β_3 = final beta coefficients; R^2 change = change in R^2—for the first step it refers to R^2; Agg_Pe = physical and verbal aggression peer evaluation at age 12.

†Significant at $p \leq .07$. **Significant at $p \leq .01$. ***Significant at $p \leq .001$.

Discussion

As previous studies have highlighted (Bandura, 1999; Bandura et al., 1996, 2001; Caprara et al., 1996), the longitudinal findings of the present study attested to the important predictive role that moral disengagement plays in the study of aggressive and violent behaviors. At the population level, for both boys and girls, the developmental model attested to the general tendency of moral disengagement to decline over time. In particular, moral disengagement decreased strongly between ages 14 and 16 and less evidently until age 20. The general decrease in moral disengagement could be specific to the developmental period from early adolescence (junior high school) to adolescence (high school), which is characterized by new challenges related to educational and social role transitions. It could reflect a change in cognitive and social structures and processes through development of the capacity to assign meaning, to anticipate outcomes, to plan actions, and to learn from social experiences the value of assigning different behaviors. Generally, over the course of maturation, individuals' social adjustment and self-regulatory abilities improve and their ability to infer the perspective of others increases (Bandura, 1991, 2004), promoting moral reasoning and moral agency (Eisenberg, 2000; Eisenberg et al., 2006). Indeed, adolescents are exposed to novel "role-taking opportunities" and moral dilemmas with personally relevant consequences to the self and to others (Hart & Atkins, 2002; Hart, Atkins, Markey, & Youniss, 2004). It may be that these refinements in skills are indispensable in preventing disengagement in "moral life" during adolescence, a time when individuals start to become more responsible and agentic in their social life than was typically possible in childhood (Hart & Carlo, 2005).

Furthermore, results attested to the significant variation in individual differences in the initial status and in the general developmental pathway of moral disengagement, supporting the importance of further exploring interindividual differences during development. At this level, the developmental trajectories model confirmed what was found at the population level and provides a more comprehensive picture of the development of moral disengagement. In fact, although distinct groups of adolescents followed different developmental trends, most adolescents (89%) decreased their level of moral disengagement during this time. As expected, the best developmental trajectory model included four different trajectory groups in accordance with the existing aggression literature (Barker et al, 2006; Brame et al., 2001; Broidy et al., 2003; Côté et al., 2002; Nagin & Tremblay, 1999; Tremblay & Nagin, 2005) and other developmental studies on aggressogenic personality factors such as irritability and hostile rumination, which have been shown to be related to moral disengagement and aggression (Caprara et al., 2007). The largest group may be considered as representing the "normative" group and started with medium levels of moral disengagement, which subsequently decreased during development. A similar trend was followed by the group called "nondisengaged," which started with lower levels of moral disengagement than the normative group. The last two smaller groups started with initial high-medium levels of moral disengagement but followed different developmental trends. The "later desister group" showed a discontinuous pattern of moral disengagement decrease. In particular, their level of moral disengagement increased from ages 14 to 16 and then strongly decreased from ages 16 to 20. In contrast, the "chronic group" showed stable levels of moral disengagement during adolescence. Although these two groups demonstrated similar physical and verbal aggression and violence profiles from ages 14 to 16; from ages 18 to 20, they showed different developmental trends: The chronic group maintained their high levels of aggression, whereas the later desister group demonstrated decreasing levels of aggression. These two groups were similar to findings of previous longitudinal studies that have identified a group of youth who show chronically high levels of antisocial and aggressive behavior during adolescence and adulthood (Brame et al., 2001; Moffitt, Caspi, Harrington, & Milne, 2002) and a group that later decrease their level of problematic behaviors during adolescence (Moffitt, 1993; Nagin, Farrington, & Moffit, 1995).

Both the later desister and chronic groups call one's attention to two possible developmental trends for adolescents that show high levels of moral disengagement at age 14. Later desister adolescents decrease their level of moral disengagement probably by improving, over the course of maturation, their cognitive and emotional skills, even if they continued to show higher levels of violence than their normative or nondisengaged peers. For chronic disengaged adolescents, moral disengagement could instead reflect a strategy of adaptation that is embedded into a system of beliefs about the self and others and leads to perceive aggression and violence as appropriate means to pursue one's own goals. Adolescents who consider aggression a morally acceptable behavior are more likely to act persistently aggressive (Crane-Ross, Tisak, & Tisak, 1998; Huesmann & Guerra, 1997; Keltikangas-Järvinen, 2001) and to avoid affective self-evaluative reactions like guilt (Bandura, 1991;

Bandura et al., 1996). Profile analysis demonstrated that at ages 14 and 16, the chronic group showed the same level of need for reparation as did the normative and the later desister groups. With increasing age, the chronic group decreased their level of need for reparation, whereas the later desister group continued to show the same level as the normative group. These results highlight how affective self-evaluative reactions were associated with different moral disengagement trends. The need for reparation associated with the negative outcomes of guilt-eliciting actions indicated that self-sanctions even existed at age 14, especially for the low disengaged group. During the course of adolescence, the chronic recourse to moral disengagement mechanisms consented people to attenuate self-sanctions and guilt feelings.

In accordance with previous findings (Bandura, 1999, 2004; Bandura et al., 1996), boys showed higher levels of moral disengagement than did girls and a lower decline from ages 14 to 16. Furthermore, adolescents in the chronic group were mostly boys, whereas most of nondisengaged group were girls. Several studies have attested that girls have higher moral emotional competence than boys (Eisenberg, 2000; Ferguson & Crowley, 1997; Lennon & Eisenberg, 1987; Mills, Pedersen, & Grusec, 1989) and lower levels of moral disengagement (Bandura, 2004; Bandura et al., 1996; Pelton et al., 2004) and aggressive behaviors (Bettencourt & Miller, 1996; Knight, Guthrie, Page, & Fabes, 2002). Furthermore, results showed that boys evaluated as aggressive by their peers already in early adolescence were more at risk to recourse to moral disengagement than girls. Moreover, both boys and girls that were never nominated by their classmates had a higher probability of being a member of both the normative and the nondisengaged groups. Thus, it is more probable that adolescents, especially boys, who have already shown aggressive behaviors in early adolescence, have more opportunities to recourse to moral disengagement.

Regarding the outcomes, the stability and change of moral disengagement significantly affected aggressive and violent behaviors. In fact, on the one hand, decreasing levels of moral disengagement at the population level were significantly linked with lower levels of physical and verbal aggression and violence in late adolescence. On the other hand, the analysis of the predictive power of moral disengagement trajectories showed that different trends were differently associated with aggressive and violent outcomes. For the nondisengaged youth (above all girls), the decline over time was negatively associated with aggressive and violent behavior. Instead, for the later desister youth, which were more likely to be at risk, the decline over time was not significantly associated with aggressive and violent outcomes. In this last case (7% of total sample), we cannot exclude the beneficial effect due to a later development of social emotional skills. For the chronic group, who showed high stable levels of moral disengagement over the course of time and the lowest levels of self-sanction for their offensive behaviors in late adolescence, the relation between moral disengagement development and aggressive and violent outcomes was particularly problematic. In fact, the results indicated that adolescents with a high probability of being a member of this group were more likely to show frequent aggressive and violent behavior in late adolescence, controlling for gender and peer nomination of physical and verbal aggression at age 12. For these youth, the stable tendency toward moral disengagement may reflect a crystallization of moral disengagement mechanisms over time that subsequently legitimatizes the recourse to aggressive and violent behaviors. The influence that moral disengagement exerted on engagement in aggression and violent episodes for these adolescence (above boys) calls one's attention to the psychological processes and social models that provide the cognitive framework within which retaliation, revenge, and violence appear appropriate and acquire legitimacy (Arsenio & Lemerise, 2004; Caprara, 1996).

In sum, the longitudinal findings of the present study indicated that: (a) most adolescents exhibited declining levels of moral disengagement over time; (b) adolescents who exhibited lower levels of moral disengagement than the normative group were typically girls, less likely to show aggression and violent acts in late adolescence and more likely to feel guilt; (c) adolescents who showed high levels of aggression in early adolescence were more likely to recourse to moral disengagement; (d) adolescents who maintained higher levels of moral disengagement were typically boys, were more likely to show frequent physical and verbal aggression and violent acts, and were less likely to feel guilt in late adolescence; and (e) adolescents who were more likely to exhibit aggression and violent acts in late adolescence were typically boys evaluated as aggressive by peers in early adolescence.

Limitations and Future Directions

The identification of developmental pathways of moral disengagement represents a first step in understanding the processes that drive some people to resort to aggression and violence more frequently than others in late adolescence. However, clarification of some issues remains necessary. First, our findings call for some caution in light of the possible limitations of the study due to the specific cultural context where the research was conducted and the use of self-report data. Second, because there are different approaches in studying moral development, future research should combine the study of moral disengagement (see Bandura, 1991) with the examination of moral reasoning and moral judgments (Kochanska & Aksan, 2006; Yau & Smetana, 2003) over the course of development. Third, future studies should examine the role of self-evaluative reactions in preventing the chronic recourse to moral disengagement mechanisms that may get crystallized over time. In this respect, different studies have highlighted the importance of studying affective reactions such as guilt, remorse, and empathy in association with moral development (Eisenberg, 2000; Hoffman, 1979; Kochanska & Aksan, 2006; Tangney et al., 2007). Finally, further studies should clarify how differences in moral disengagement are associated with emotional, cognitive, and social processes implied in moral self-regulation (Hart & Carlo, 2005; Turiel, 2006) and how individual skills and environmental factors could affect both individual differences and aggressive and violent outcomes. In this regard, we believe that it is necessary to combine the systematic study of individual differences with the study of external situational variables within a longitudinal perspective. The study of both internal mechanisms and relevant environmental variables that operate at stimulus, social, contextual, structural, and cultural levels to influence individual, and group behaviors are necessary in designing appropriate interventions aimed at preventing aggressive and violent outcomes.

References

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin, 103,* 411–423.

Andrus, B. C. (1969). *The infamous of Nuremberg*. London: Fravin.

Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 243–277). Mahwah, NJ: Erlbaum.

Archer, J. (2004). Sex differences in aggression in real-world settings: A meta-analytic review. *Review of General Psychology, 8,* 291–322.

Arsenio, W., & Lemerise, E. (2004). Aggression and moral development: Integrating the social information processing and moral domain models. *Child Development, 75,* 987–1002.

Bandalos, D. L., & Finney, S. J. (2001). Item parceling issues in structural equation modeling. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: New developments and techniques* (pp. 269–296). Mahwah, NJ: Erlbaum .

Bandura, A. (1986). *Social foundation of thought and action: A social cognitive theory.* Englewood Cliffs, NJ: Prentice Hall.

Bandura, A. (1990). Mechanism of moral disengagement. In W. Reich (Ed.), *Origin of terrorism: Psychologies, ideologies, theologies, states of mind* (pp. 161–191). Cambridge, UK: Cambridge University Press.

Bandura, A. (1991). Social cognitive theory of moral thought and action. In W. M. Kurtines & J. L. Gewirtz (Eds.), *Handbook of moral behaviour and development* (Vol. I, pp. 45–103). Hillsdale, NJ: Erlbaum.

Bandura, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review, 3,* 193–209.

Bandura, A. (2001). *Social cognitive theory: An agentic perspective. Annual review of psychology* (Vol. 52, pp. 1–26). Palo Alto, CA: Annual Reviews.

Bandura, A. (2004). Selective exercise of moral agency. In T. A. Thorkildsen & H. J. Walberg (Eds.), *Nurturing morality* (pp. 37–57). Boston: Kluwer Academic.

Bandura, A. (2006). Toward a psychology of human agency. *Perspectives on Psychological Science, 1,* 164–180.

Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (1996). Mechanism of moral disengagement in the exercise of moral agency. *Journal of Personality and Social Psychology, 71,* 364–374.

Bandura, A., Caprara, G. V., Barbaranelli, C., Pastorelli, C., & Regalia, C. (2001). Sociocognitive self-regulatory mechanisms governing transgressive behavior. *Journal of Personality and Social Psychology, 80,* 125–135.

Bandura, A., Caprara, G. V., & Zsolnai, L. (2000). Corporate transgressions through moral disengagement. *Journal of Human Values, 6,* 57–64.

Bandura, A., Underwood, B., & Fromson, M. E. (1975). Disinhibition of aggression through diffusion of responsibility and dehumanizaion of victims. *Journal of Research in Personality, 9,* 253–269.

Bandura, A., & Walters, R. H. (1959). *Adolescent aggression: A study of the influence of child-training practices and family interrelationships.* Oxford, UK: Ronald Press.

Barker, E. D., Tremblay, R. E., Nagin, D. S., Vitaro, F., & Lacourse, E. (2006). Development of male proactive and reactive physical aggression during adolescence. *Journal of Child Psychology and Psychiatry, 47,* 783–790.

Bentler, P. M. (1989). *EQS Structural Equations Program Manual. Manual for Version 3.0.* Los Angeles: BMDP Statistical Software.

Bergman, R. (2002). Why be moral? A conceptual model from developmental psychology. *Human Development, 45,* 104–124.

Bettencourt, A. B., & Miller, N. (1996). Gender differences in aggression as a function of provocation: A meta-analysis. *Psychological Bulletin, 119,* 422–447.

Blasi, A. (1984). Moral identity: Its role in moral functioning. In W. W. Kurtines & J. L. Gewirtz (Eds.), *Morality, moral behavior, and moral development* (pp. 128–139). New York: Wiley.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: Wiley.

Bollinger, D. (1982). *Language: The loaded weapon.* London: Longman.

Brame, B., Nagin, D. S., & Tremblay, R. E. (2001). Developmental trajectories of physical aggression from school entry to late adolescence. *Journal of Child Psychology and Psychiatry, 42*, 503–512.

Brock, T. C., & Buss, A. H. (1962). Dissonance, aggression, and evaluation of pain. *Journal of Abnormal and Social Psychology, 65*, 197–202.

Brock, T. C., & Buss, A. H. (1964). Effects of justification for aggression and communication with the victim on postaggression dissonance. *Journal of Abnormal and Social Psychology, 68*, 403–412.

Broidy, L. M., Nagin, D. S., Tremblay, R. E., Bates, J. E., Brame, B., Dodge, K. A., et al. (2003). Developmental trajectories of childhood disruptive behaviors and adolescent delinquency: A six-site, cross-national study. *Developmental Psychology, 39*, 222–245.

Caprara, G. V. (1996). Structures and processes in personality psychology. *European Psychologist, 1*, 14–26.

Caprara, G. V., Bandura, A., Barbaranelli, C., & Vicino, S. (1996). La Misura del Disimpegno morale [The assessment of moral disengagement]. *Rassegna di Psicologia, 13*, 93–105.

Caprara, G. V., Barbaranelli, C., Pastorelli, C., Cermak, I., & Rosza, S. (2001). Facing guilt: Role of negative affectivity, need for reparation, and fear of punishment in leading to prosocial behaviour and aggression. *European Journal of Personality, 15*, 219–237.

Caprara, G. V., & Capanna, C. (2005). Moral-civic disengagement and values. *Ricerche di Psicologia, 28*(1), 67–83.

Caprara, G. V., Manzi, J., & Perugini, M. (1992). Investigating guilt in relation to emotionality and aggression. *Personality and Individual Differences, 13*, 519–532.

Caprara, G. V., Mazzotti, E., & Prezza, M. (1990). Una scala per la misura dell'atteggiamento verso la violenza [A scale for the measurement of attitude toward violence]. *Giornale Italiano di Psicologia, 17*, 107–120.

Caprara, G. V., Paciello, M., Gerbino, M., & Cugini, C. (2007). Individual differences conducive to aggression and violence: Trajectories and correlates of irritability and hostile rumination through adolescence. *Aggressive Behavior, 33*, 359–374.

Caprara, G. V., & Pastorelli, C. (1993). Early emotional instability. Prosocial behavior and aggression: Some methodological aspects. *European Journal of Personality, 7*, 19–36.

Carlo, G., Eisenberg, N., & Knight, G. P. (1992). An objective measure of adolescents' prosocial moral reasoning. *Journal of Research on Adolescence, 2*, 331–349.

Cattell, R. B., & Vogelmann, S. (1977). A comprehensive trial of the scree and KG criteria for determining the number of factors. *Multivariate Behavioral Research, 12*, 289–325.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*, 233–255.

Cohen, D., & Nisbett, R. E. (1994). Self-protection and the culture of honor: Explaining southern violence. *Personality and Social Psychology Bulletin, 20*, 551–567.

Côté, S., Tremblay, R. E., Nagin, D. S., Zoccolillo, M., & Vitaro, F. (2002). Childhood behavioral profiles leading to adolescent conduct disorder: Risk trajectories for boys and girls. *Journal of the American Academy of Child and Adolescent Psychiatry, 41*, 1086–1094.

Crane-Ross, D., Tisak, M. S., & Tisak, J. (1998). Aggression and conventional rule violation among adolescents: Social-reasoning predictors of social behavior. *Aggressive Behavior, 24*, 347–365.

Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information-processing mechanism in children's social adjustment. *Psychological Bulletin, 115*, 74–101.

Darley, J. M., Klosson, E. C., & Zanna, M. P. (1978). Intentions and their contexts in the moral judgments of children and adults. *Child Development, 49*, 66–74.

Diener, E. (1977). Deindividuation: Causes and consequences. *Social Behavior and Personality, 5*, 145–156.

Diener, E., Dineen, J., Endresen, K., Beaman, A. L., & Fraser, S. C. (1975). Effects of altered responsibility, cognitive set, and modeling on physical aggression and deindividuation. *Journal of Personality and Social Psychology, 31*, 328–337.

Dodge, K. A., & Crick, N. R. (1990). Social information-processing bases of aggressive behavior in children. *Personality and Social Psychology Bulletin, 16*, 8–22.

Duncan, S. C., & Duncan, T. E. (1994). Modeling incomplete longitudinal substance use data using latent variable growth curve methodology. *Multivariate Behavioral Research, 29*, 313–338.

Duncan, T. E., & Duncan, S. C. (1995). Modeling the processes of development via latent growth curve methodology. *Structural Equation Modeling, 2*, 187–213.

Duncan, T. E., Duncan, S. C., & Hops, H. (1994). The effects of family cohesiveness and peer encouragement on the development of adolescence alcohol use: A cohort-sequential approach to the analysis of longitudinal data. *Journal of Studies on Alcohol, 55*, 588–599.

Eisenberg, N. (2000). Emotion, regulation, and moral development. *Annual Review of Psychology, 51*, 665–697.

Eisenberg, N., Fabes, R. A., & Spinrad, T. L. (2006). *Prosocial development.* Hoboken, NJ: John Wiley & Sons.

Elkin, F., & Westley, W. A. (1955). The myth of adolescent culture. *American Sociological Review, 20*, 680–684.

Elliott, D. S., & Rhinehart, M. (1995). *Moral disengagement, delinquent peers and delinquent behaviour.* Unpublished manuscript, University of Colorado, Institute of Behavioral Science.

Emmons, R. A., & Diener, E. (1986). Situation selection as a moderator of response consistency and stability. *Journal of Personality and Social Psychology, 51*, 1013–1019.

Ferguson, T. J., & Crowley, S. L. (1997). Gender differences in the organization of guilt and shame. *Sex Roles, 37*, 19–43.

Ferguson, T. J., & Rule, B. G. (1983). An attributional perspective on anger and aggression. In R. G. Geen & E. I. Donnerstein (Eds.), *Aggression: Theoretical and empirical reviews* (Vol. 1, pp. 41–74). New York: Academic Press.

Fix, E., Hodges, J. L., & Lehmann, E. L. (1959). The restricted chi-square test. In U. Grenanader (Ed.), *Probability and statistics: The herald Cramér volume* (pp. 92–97). New York: Wiley.

Gibbs, J. C., Potter, G. B., Barriga, A. Q., & Liau, A. K. (1996). Developing the helping skills and prosocial motivation of aggressive adolescents in peer group programs. *Aggression and Violent Behavior, 1*, 283–305.

Gini, G. (2006). Social cognition and moral cognition in bullying: What's wrong? *Aggressive Behavior, 32*, 528–539.

Hanson, W. B., Tobler, N. S., & Graham, J. W. (1990). Attrition in substance abuse prevention research: A meta analysis of 85 longitudinally followed cohorts. *Evaluation Review, 14*, 677–685.

Hardy, S. A., & Carlo, G. (2005). Identity as a source of moral motivation. *Human Development, 48*, 232–256.

Haritos-Fatouros, M. (1988). The official torturer: A learning model for obedience to the authority of violence. *Journal of Applied Social Psychology, 18*, 1107–1120.

Hart, D., & Atkins, R. (2002). Civic competence in urban youth. *Applied Developmental Science, 6*, 227–236.

Hart, D., Atkins, R., Markey, P., & Youniss, J. (2004). Youth bulges in communities: The effects of age structure on adolescent civic knowledge and civic participation. *Psychological Science, 15*, 591–597.

Hart, D., & Carlo, G. (2005). Moral development in adolescence. *Journal of Research on Adolescence, 15*, 223–233.

Helwig, C. C., & Turiel, E. (2003). Children's social and moral reasoning. In P. K. Smith & C. H. Hart (Eds.), *Blackwell handbook of childhood social development* (pp. 476–490). Malden, MA: Blackwell.

Hoeksma, J. B., & Kelderman, H. (2006). On growth curves and mixture models. *Infant and Child Development, 15*, 627–634.

Hoffman, M. L. (1979). Development of moral thought, feeling, and behavior. *American Psychologist, 34*, 958–966.

Hoyle, R. H. (1995). *Structural equation modeling: Concepts, issues, and applications*. Thousand Oaks, CA: Sage.

Hoyle, R. H., & Smith, G. T. (1994). Formulating clinical research hypotheses as structural equation models: A conceptual overview. *Journal of Consulting and Clinical Psychology, 62*, 429–440.

Huesmann, L. R. (1988). An information processing model for the development of aggression. *Aggressive Behavior, 14*, 13–24.

Huesmann, L. R. (1998). *The role of social information processing and cognitive schema in the acquisition and maintenance of habitual aggressive behavior.* San Diego, CA: Academic Press.

Huesmann, L. R., & Guerra, N. G. (1997). Children's normative beliefs about aggression and aggressive behavior. *Journal of Personality and Social Psychology, 72*, 408–419.

Istituto Italiano di Statistica. (2002). *Annuario statistico Italiano 2000.* [Italian yearbook of statistics 2002]. Rome, Italy: Author.

Jones, B. L., & Nagin, D. S. (2007). Advances in group-based trajectory modeling and an SAS procedure for estimating them. *Sociological Methods and Research, 35*, 542–571.

Jones, B. L., Nagin, D., & Roeder, K. (2001). A SAS procedure based on mixture models for estimating developmental trajectories. *Sociological Methods and Research, 29*, 374–393.

Jöreskog, K. G., & Sörbom, D. (1984). *LISREL VI: Analysis of linear structural relationships by method of maximum likelihood.* Chicago: National Educational Resources.

Keen, S. (1986). *Faces of the enemy.* San Francisco: Harper & Row.

Kelman, H. C. (1973). Violence without moral restraint: Reflections on the dehumanization of victims and victimizers. *Journal of Social Issues, 29*, 25–61.

Kelman, H. C., & Hamilton, V. L. (1989). *Crimes of obedience: Toward a social psychology of authority and responsibility.* New Haven, CT: Yale University Press.

Keltikangas-Järvinen, L. (2001). Aggressive behaviour and social problem-solving strategies: A review of the findings of a seven-year follow-up from childhood to late adolescence. *Criminal Behaviour and Mental Health, 11*(4), 236–250.

Killen, M., & Smetana, J. G. (2006). *Handbook of moral development*. Mahwah, NJ: Erlbaum.

Klass, E. T. (1978). Psychological effects of immoral actions: The experimental evidence. *Psychological Bulletin, 85*, 756–771.

Knight, G. P., Guthrie, I. K., Page, M. C., & Fabes, R. A. (2002). Emotional arousal and gender differences in aggression: A meta-analysis. *Aggressive Behavior, 28*, 366–393.

Kochanska, G., & Aksan, N. (2006). Children's conscience and self-regulation. *Journal of Personality, 74*, 1587–1617.

Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In D. A. Goslin (Ed.), *Handbook of socialization theory and research.* Chicago: McNally.

Kohlberg, L., & Candee, D. (1984). The relationship of moral judgement to moral action. In W. M. Kurtines & J. Gewirtz (Eds.), *Morality, moral behavior, and moral development* (pp. 52–73). New York: Wiley.

Krettenauer, T. (2004). Metaethical cognition and epistemic reasoning development in adolescence. *International Journal of Behavioral Development, 28*, 461–470.

Kwak, K., & Bandura, A. (1997). *Role of perceived self-efficacy and moral disengagement in antisocial conduct*. Unpublished manuscript, Osan College, Seoul, Korea.

Lennon, R., & Eisenberg, N. (1987). Gender and age differences in empathy and sympathy. In N. Eisenberg & J. Strayer (Eds.), *Empathy and its development* (pp. 195–217). New York: Cambridge University Press.

Little, R. J. (1995). Modeling the dropout mechanism in repeated-measures studies. *Journal of the American Statistical Association, 90*, 1112–1121.

Loeber, R. (1991). Antisocial behavior: More enduring than changeable? *Journal of the American Academy of Child and Adolescent Psychiatry, 30*, 393–397.

Loeber, R., & Hay, D. (1997). Key issues in the development of aggression and violence from childhood to early adulthood. *Annual Review of Psychology, 48*, 371–410.

Loeber, R., & Stouthamer-Loeber, M. (1998). Development of juvenile aggression and violence: Some common misconceptions and controversies. *American Psychologist, 53*, 242–259.

Lutz, W. D. (1987). Language, appearance, and reality: Doublespeak in 1984. *ETC, 44*, 382–391.

Martinez, M. L., Black, M., & Starr, R. H. (2002). Factorial structure of the perceived neighbourhood scale (PNS): A test of longitudinal invariance. *Journal of Community Psychology, 30*, 23–43.

McArdle, J. J. (1988). Dynamic but structural equation modelling of repeated measures data. In R. B. Cattel & J. Nesselroade (Eds.), *Handbook of multivariate experimental psychology* (2nd ed., pp. 561–614). New York: Plenum Press.

McArdle, J. J., & Anderson, E. R. (1989). Latent growth model for research on aging. In L. E. Biren & K. W. Schaie (Eds.), *The handbook of the psychology of aging* (3rd ed., pp. 21–44). San Diego: Academic Press.

McArdle, J. J., & Hamagami, F. (1991). Modeling incomplete longitudinal and cross-sectional data using latent growth structural models. In L. M. Collins & J. C. Horn (Eds.), *Best methods for the analysis of change* (pp. 276–304). Washington, DC: American Psychological Association.

Menesini, E., Sanchez, V., Fonzi, A., Ortega, R., Costabile, A., & Lo Feudo, G. (2003). Moral emotions and bullying: A cross-national comparison of differences between bullies, victims and outsiders. *Aggressive Behavior, 29*, 515–530.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*, 525–543.

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika, 55*, 107–122.

Milgram, S. (1974). *Obedience to authority: An experimental view.* New York: Harper & Row.

Mills, R. S., Pedersen, J., & Grusec, J. E. (1989). Sex differences in reasoning and emotion about altruism. *Sex Roles, 20*, 603–621.

Moffitt, T. E. (1993). Adolescence-limited and life-course-persistent antisocial behavior: A developmental taxonomy. *Psychological Review, 100*, 674–701.

Moffitt, T. E., Caspi, A., Harrington, H. L., & Milne, B. J. (2002). Males on the life-course persistent and adolescence-limited antisocial pathways: Follow-up at age 26. *Development and Psychopathology, 14*, 179–207.

Motl, R. W., Dishman, R. K., Birnbaum, A. S., & Lytle, L. A. (2005). Longitudinal invariance of the center for epidemiological studies-depression scale among girls and boys in middle school. *Educational and Psychological Measurement, 65*, 90–108.

Muthén, B. O., & Khoo, S. (1998). Longitudinal studies of achievement growth using latent variable modeling. *Learning and Individual Differences, 10*, 73–101.

Muthén, B., & Shedden, K. (1999). Finite mixture modelling with mixture outcomes using the EM algorithm. *Biometrics, 6*, 463–469.

Nagin, D. S. (1999). Analyzing developmental trajectories: A semiparametric, group-based approach. *Psychological Methods, 4*, 139–157.

Nagin, D. S. (2005). *Group-based modeling of development*. Cambridge, MA: Harvard University Press.

Nagin, D. S., Farrington, D. P., & Moffit, T. E. (1995). Life-course trajectories of different types of offenders. *Criminology, 33*, 111–139.

Nagin, D., & Tremblay, R. E. (1999). Trajectories of boys' physical aggression, opposition, and hyperactivity on the path to physically violent and nonviolent juvenile delinquency. *Child Development, 70*, 1181–1196.

Pelton, J., Gound, M., Forehand, R., & Brody, G. (2004). The moral disengagement scale: Extension with an American minority sample. *Journal of Psychopathology and Behavioral Assessment, 26*, 31–39.

Pepler, D., Jiang, D., Craig, W., & Connolly, J. (2008). Developmental trajectories of bullying and associated factors. *Child Development, 79*, 325–338.

Pitts, S. C., West, S. G., & Tein, J. (1996). Longitudinal measurement models in evaluation research: Examining stability and change. *Evaluation and Program Planning, 19*, 333–350.

Rapoport, D. C., & Alexander, Y. (1982). *The morality of terrorism: Religious and secular justification.* Elmsford, NY: Pergamon Press.

Raykov, T. (2004). Behavioral scale reliability and measurement invariance evaluation using latent variable modeling. *Behavior Therapy, 35*, 299–331.

Raykov, T., & Marcoulides, G. A. (2006). *A first course in structural equation modeling* (2nd ed.). Mahwah, NJ: Erlbaum.

Reich, W. (1990). *Origins of terrorism: Psychologies, ideologies, theologies, states of mind.* Cambridge, UK: Cambridge University Press.

Sampson, R. J., & Laub, J. H. (2003). Life-course desisters? Trajectories of crime among delinquent boys followed to age 70. *Criminology, 41*, 555–592.

Sanford, N., & Comstock, C. (1971). *Sanction for evil.* San Francisco: Jossey-Bass.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological methods, 7*, 147–177.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*, 461–464.

Steiger, J. H., & Lind, J. C. (1980, May). *Statistically based tests for the number of common factors*. Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.

Stoolmiller, M. (1994). Antisocial behavior, delinquent peer association, and unsupervised wandering for boys: Growth and change from childhood to early adolescence. *Multivariate Behavioral Research, 29*, 263–288.

Swanson, L. H., & Hill, G. (1993). Metacognitive aspects of moral reasoning and behavior. *Adolescence, 28*, 711–736.

Tabachnick, B. G., & Fidell, L. S. (1989). *Using multivariate statistics*. New York: Harper & Row.

Tangney, J. P., Stuewig, J., & Mashek, D. J. (2007). Moral emotions and moral behavior. *Annual Review of Psychology, 58*, 345–372.

Thurstone, L. L. (1947). *Multiple-factor analysis*. Chicago: University of Chicago Press.

Tremblay, R. E. (2000). The development of aggressive behavior during childhood: What have we learned in the past century? *International Journal of Behavioral Development, 24*, 129–141.

Tremblay, R. E., & Nagin, D. S. (2005). The developmental origins of physical aggression in humans. In R. E. Tremblay, W. W. Hartup, & J. Archer (Eds.), *Developmental origins of aggression* (pp. 83–106). New York: Guilford Press.

Turiel, E. (1978). The development of concepts of social structure: Social convention. In J. Glick & K. A. Clarke-Stewart (Eds.), *The development of social understanding*. New York: Gardner Press.

Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge, UK: Cambridge University Press.

Turiel, E. (2006). *Thought, emotions, and social interactional processes in moral development*. Mahwah, NJ: Erlbaum.

Weiner, B. (1986). *An attributional theory of motivation and emotion*. New York: Springer-Verlag.

West, S. G., Finch, J. F., & Curran, P. J. (1995). *Structural equation models with nonnormal variables: Problems and remedies*. Thousand Oaks, CA: Sage Publications.

Widaman, K. F., & Reise, S. P. (1997). *Exploring the measurement invariance of psychological instruments: Applications in the substance use domain*. Washington, DC: American Psychological Association.

Yau, J., & Smetana, J. G. (2003). Conceptions of moral, social-conventional, and personal events among Chinese preschoolers in Hong Kong. *Child Development, 74*, 647–658.

Zimbardo, P. G. (1995). The psychology of evil: A situationist perspective on recruiting good people to engage in anti-social acts. *Research in Social Psychology, 11*, 125–133.

Zimbardo, P. G. (2004). A situationist perspective on the psychology of evil: Understanding how good people are transformed into perpetrators. In A. G. Miller (Ed.), *The social psychology of good and evil* (pp. 21–50). New York: Guilford Press.

Zimbardo, P. G. (2006). The psychology of power: To the person? To the situation? To the system?. In D. L. Rhode (Ed.), *Moral leadership: The theory and practice of power, judgment and policy. Warren Bennis Signatures Series* (pp. 129–157). San Francisco: Jossey-Bass.

Appendix

Moral Disengagement Scale

1. It is alright to fly off the handle to protect your friends.
2. Slapping and shoving someone is just a way of joking.
3. Damaging some property is no big deal when you consider that others are beating people up.
4. A member of a group should not be blamed for trouble the group causes.
5. If youth are living under bad conditions in their neighbourhood they cannot be blamed for behaving aggressively.
6. It is not serious to tell small lies because they don't hurt anybody.
7. Some people deserve to be treated like animals.
8. If people fight and misbehave in school or at work it is their teacher's /superior's fault.
9. It is alright to beat someone who bad mouths your family.
10. To hit obnoxious friends is just giving them "a lesson."
11. Stealing some money is not too serious compared to those who steal a lot of money.
12. A person who only suggests breaking rules should not be blamed if others go ahead and do it.
13. If youth are not disciplined at home they should not be blamed for misbehaving.
14. People do not mind being teased because it shows interest in them.
15. It is okay to treat somebody badly who behaved like a "worm."
16. If people are careless about where they leave their things it is their own fault if they get stolen.
17. It is alright to fight when your group's honour is threatened.
18. Taking someone's motorcycle or car without their permission is just "borrowing it."
19. It is not serious to insult a friend because beating him/her up is worse.
20. If a group decides together to do something harmful it is unfair to blame a single member of the group for it.
21. Youths cannot be blamed for using bad words when all their friends do it.

22. Teasing someone does not really hurt him/her.
23. Someone who is detestable does not deserve to be treated like a human being.
24. People who get mistreated usually do things that deserve it.
25. It is alright to lie to keep your friends out of trouble.
26. It is not a bad thing to "get drunk" once in a while.
27. Compared to the illegal things people do, taking some things from a store without paying for them is not very serious.
28. It is unfair to blame a single person who had only a small part in the harm caused by a group.
29. Youth cannot be blamed for misbehaving if their friends pressured them to do it.
30. Insults among peers do not hurt anyone.
31. Some people have to be treated roughly because they lack feelings that can be hurt.
32. Youths are not at fault for misbehaving if their parents are too restrictive (severe, and they don't allow them any freedom).

The following items correspond to the various mechanisms of moral disengagement. *Moral justification*: 1, 9, 17, 25. *Euphemistic language*: 2, 10, 18, 26. *Advantageous comparison*: 3, 11, 19, 27. *Displacement of responsibility*: 5, 13, 21, 29. *Diffusion of responsibility*: 4, 12, 20, 28. *Distorting consequences*: 6, 14, 22, 30. *Attribution of blame*: 8, 16, 24, 32. *Dehumanization*: 7, 15, 23, 31.

Child Development, September/October 2008, Volume 79, Number 5, Pages 1310 – 1328

Arguing on the Computer: A Microgenetic Study of Developing Argument Skills in a Computer-Supported Environment

Deanna Kuhn, Wendy Goh, Kalypso Iordanou, and David Shaenfield
Columbia University

We report a study of a class of 28 sixth graders engaged in an extended computer-supported argumentive discourse activity. Participants collaborated with a same-side peer in arguing against successive pairs of peers on the opposing side of an issue. Meta-level awareness was facilitated by conducting the dialogs via instant messaging software, which made available a transcript of the dialog that was used in additional reflective activities. In the course of dialogs on 3 successive topics, participants showed significant gains in meta-level communications about the discourse, reflecting at least implicit understanding of its goals, as well as in the strategic moves that constituted the discourse. The latter advances remained evident when the social support of a same-side partner was withdrawn.

The central place of argument skills in the development of higher order thinking is increasingly being recognized. Indeed, the construction and evaluation of arguments constitute a major component of the College Learning Assessment (Hersh, 2005), designed to assess the cognitive development that occurs during the college years. Until recently, research on argument skills has been devoted almost entirely to individuals' written or verbal arguments in support of a claim. Studies have been consistent in revealing skills that are at best modest among children and adolescents and show little improvement thereafter (Brem & Rips, 2000; Glassner, Weinstock, & Neuman, 2005; Kuhn, 1991; Means & Voss, 1996; Perkins, 1985; Stanovich, 1999; Voss & Means, 1991; Weinstock, Neuman, & Tabak, 2004).

The work presented here is part of a growing body of research that examines the process of argumentation, or argumentive discourse, in older children and adolescents (Voss, 2001). As elaborated elsewhere (Graff, 2003; Kuhn, 2005), such discourse is of interest for its alleged role in facilitating the development of individual (nonsocial) argument skills. We agree with Graff's (2003) view that dialogic argument provides the "missing interlocutor" that lends purpose to and thereby facilitates development of individual argument skills.

Although competence in individual argument skills is clearly critical, most notably in academic contexts, the development of dialogic argumentation skills is important in its own right. What needs to develop in the domain of argumentive discourse skill and how does this development occur? The present work is focused on these questions, although we include an assessment of the extent to which gains observed in argumentive discourse affect performance on a measure of individual argument.

Dialogic argumentation has long been of broad interest to developmentalists, particularly those who regard social collaboration as central to cognitive development (Damon, 1984; Moshman, 2005; Smetana, Killen, & Turiel, 1991; Wainryb, Shaw, Laupa, & Smith, 2001). The focus of more recent, educationally oriented research on argumentive discourse has been on devising scaffolds (typically computer software) to support the argumentation process (Andriessen, 2006; Andriessen, Baker, & Suthers, 2004; Bell & Linn, 2000; Chinn, 2006; Clark & Sampson, 2005; Glassner & Schwarz, 2005; Nussbaum, 2005; Wiley & Voss, 1999; Zohar & Nemet, 2002). The work of Anderson and colleagues (Anderson et al., 2001; Reznitskaya et al., 2001) and, following them, Nussbaum (2003) is an exception in examining naturally occurring discourse in small groups of elementary-school-age children and seeking to identify the roots of more sophisticated argumentation skills as they emerge in such conversation and spread from one child to others.

Dialogic argument is a skill that has the advantage of early roots in everyday conversation, and yet, the demands of skilled argumentive discourse are considerable, with previous research indicating that few individuals attain proficiency during the second decade of life (Felton & Kuhn, 2001; Kuhn & Franklin, 2006; Kuhn & Udell, 2003; Udell, 2007). At the same time that one is processing and evaluating input from the conversational partner, one must be formulating

---

Correspondence concerning this article should be addressed to Deanna Kuhn, Box 119, Department of Human Development, Teachers College Columbia University, New York, NY 10027. Electronic mail may be sent to dk100@columbia.edu.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7905-0007

an effective response that meets discourse goals. According to Walton (1989), skilled argumentation has two goals. One is to secure commitments from the opponent that can be used to support one's own argument. The other is to undermine the opponent's position by identifying and challenging weaknesses in his or her argument. Both these goals, note, require attention to the opponent's claims.

In earlier work, we have found that young adolescents concentrate their efforts on exposition of their own claims to the neglect of attending to the opponent's claims and attempting to weaken their force (Felton, 2004; Felton & Kuhn, 2001; Kuhn & Udell, 2003; Udell, 2007). Yet, they are able to attend to the other's argument, we have found, and even generate an argument against it, when explicitly instructed to do so (Kuhn & Udell, 2007). The challenge, then, may be less one of executing the skill (of addressing the opposing position) than it is one of recognizing the need to do so. This recognition goes to the very heart of argument. If the opponent's argument for his or her position is not relevant, the process through which one may claim to attain victory over the opponent cannot be regarded as one of argument.

In the present work, we focus on how young adolescents develop better meta-level understanding of the goals of argument, as well as procedural skill in implementing these goals in discourse. Our prior work indicates that dense engagement over time in argumentive discourse with a series of peers does lead to advancements in procedural skill, and we seek to corroborate this finding in the present study. Our focus here, however, is on the development of meta-level understanding about argumentive discourse and its goals. Specifically, does this meta-level understanding also develop and support skill development at the procedural level, that is, in the strategies involved in the actual conduct of the argumentation?

Most generally, one must see a point to argument if one is to engage in it effectively (Kuhn, 1991, 2005). Specifically, we claim, one must understand the goal of argumentation as one of engaging one another's claims and undertaking to weaken them, as well as seeking acceptance of one's own claims—the two goals noted earlier identified by Walton (1989). Progress with respect to this meta-level development, we propose, is likely to support progress in the procedural aspect as well.

The form of dense engagement in peer discourse devised for this study was designed with the objective of supporting to the extent possible the meta-level, as well as strategic, development to which we have referred. Beyond dense engagement in discourse itself, which has been found effective with various age groups in our own previous studies (Felton, 2004; Felton & Kuhn, 2001; Kuhn, Shaw, & Felton, 1997; Kuhn & Udell, 2003; Udell, 2007) and those of several other researchers (Anderson et al., 2001; Chinn, 2006; Nussbaum, 2003; Nussbaum & Kardash, 2005; Nussbaum & Sinatra, 2003; Reznitskaya et al., 2001; Schwarz, Newman, & Biezuner, 2000), the general strategy we thought might be most effective is to heighten participants' awareness of the discourse. This meta-level awareness regarding the discourse seems a necessary condition to formulating, and ultimately implementing, goals with respect to it.

We employ three techniques designed to heighten and support this awareness. First, we ask participants to work in pairs. The two members of the pair hold the same position on an issue and they collaborate in arguments against a series of opposing pairs who hold the opposite position. Each communication to the opposing dyad is made only when the pair has agreed what it is to be. Having to deliberate regarding each argumentive move, we hypothesized, would not only generate more thoughtful contributions to the discourse but also make the collaborative pair more aware and reflective regarding this discourse.

The second technique, following on the study by Felton (2004), is to integrate an explicit reflective activity into the series of dialogs. In the present study, we do this by making available to participants some of their own dialog transcripts as the basis for this activity. Support for these design features comes from the study by Felton; although adolescents individually engaged in argument with an opposing peer, two peer advisors observed and later discussed possible improvements with their partners—a condition Felton found enhanced dialogic argument skills relative to a condition in which participants argued without peer advisors. In the present study, we engage peers in collaboratively planning and implementing the exchange itself, not only reflecting on it.

The third technique we employ here is to conduct these dialogs via computer-supported instant messaging (IM) software, a method we have not tried in previous work and the one that constitutes the major innovation of the present study. In addition to capitalizing on young teens' familiarity with the medium, this technique has the benefit of providing an immediately available and permanent record of the discourse, in striking contrast to the conditions of real-time verbal discourse, where the contents of each contribution to the dialog immediately disappear as soon as they are spoken.

The present study extends over an entire school year. We begin with assessments of individuals' argument skills, both in generating an individual written argument in support of their position on a social issue and in engaging in an electronically conducted argument with a peer who holds an opposing view on the topic. These assessments are repeated at the end of the year, allowing us to assess improvement in both individual and dialogic argument skills. A focus of our work, however, is microgenetic analysis (Kuhn, 1995; Miller & Coyle, 1999; Siegler, 2006) of the period of intervention itself. During this time, we observe participants developing meta-level understanding of argumentation, as well as developing strategic skill in executing it, as each participant and a partner engage in electronic discourse with an opposing pair on a series of new topics.

Method

Participants

The 28 participants consisted of the entire sixth grade at a university-affiliated independent school in a large urban setting. The school has a unique population in that 50% of the school's slots are reserved for children of university faculty and high-level administrators, whereas the remaining 50% of slots are filled by children from the surrounding low- to middle-income community chosen by lottery and receiving full financial support. The school was in only its 2nd year of operation, the year that the work described here took place. Hence, students had come from a wide range of previous schools and were of diverse academic, as well as ethnic, racial, and socioeconomic, backgrounds. Roughly half were identifiable as racial or ethnic minorities (largely Hispanic and African American). All were 11 or 12 years of age; 16 were girls and 12 were boys.

Participants had become familiar with the school's innovative curriculum by the time our project began. The school features an integrated curriculum across all subjects and has a strong technology emphasis, with Smartboards in each classroom and students each allocated an individual laptop computer used daily in their schoolwork.

Initial Assessment

Capital punishment (CP) was the topic on which initial and final assessments of both individual and dialogic argument were based.

Individual argument. Each participant's initial position on CP was assessed using a 13-point Likert scale. The middle position on the scale read, "I have mixed or undecided opinions about capital punishment." Each position to one side or the other of this middle point represents a position progressively more extreme in support of or against CP. The Point 1 position from the center on the "support" end of the scale, for example, read, "I am somewhat in favor of CP but I'm not sure," whereas the end position read, "I am totally in favor of CP. I will never change my opinion and I cannot imagine how someone could have a different opinion." Prior to asking participants to check a box that best represented their view, one of the adult coaches defined CP for the class ("administering the death penalty for serious crimes such as murder") and explained the scale.

Completion of the scale instrument was followed by a written assessment instrument. After again eliciting the participant's position as for, against, or undecided, the following questions were posed (with space after each for the participant to respond):

1. Why did you choose this position? In other words, what are your reasons for your position?
2. What are some other possible reasons for this position? These reasons can be your own reasons or other people's reasons who have the same position as you.
3. Can you imagine someone having another position that is different from your own? What would it be?
4. What might be some reasons for having this different position?
5. Can you think of any other reasons to be for or against CP?

Dialogic argument. The 8 participants who indicated they were undecided on the scale gave reasons on both sides of the issue in the questionnaire. These 8 were grouped with the 6 participants who chose a scale position supportive of CP. The remaining 14 participants, who indicated they were opposed to CP on both the scale and the questionnaire, formed a second group. The week following the individual assessment, each of the participants from Group 1 was paired with a participant from Group 2, and each of these pairs was asked to participate in a dialog with one another on the topic of CP.

To conduct these dialogs, all Group 1 members were situated in one classroom and all Group 2 members in an adjacent classroom. Dialogs were conducted via *FirstClass*© IM software, installed on each participant's individual laptop computer. Partners were instructed to conduct a dialog via IM. Their goal, they were instructed, was to try to convince their partners that their position was the better one. These dialogs lasted from 10 to 20 min; they ended when participants indicated they had finished or when no more time was available. The software saved the dialog for analysis.

Intervention: Collaborative Dialogic Argument Plus Reflection

The argumentation activity took place in three identical phases, each devoted to a different topic. The first of the three phases, described here in detail for illustration, comprised thirteen 40-min classroom sessions occurring twice per week. The activity was introduced to students as a class in learning to become good arguers. The class project was to focus on ColumbiaTown, a new town being formed in the middle United States. The governing board of the new town had to establish rules and laws for the new town, and in particular, the participants' task would be to focus on the new school that was being formed to serve the town's children. The first topic was to be whether attendance at school should be mandatory for all children in the town. Specifically, the following dilemma was introduced:

> The Costa family has moved to the edge of town from far away Greece with their 11-year-old son Nick. Nick was a good student and soccer player back home in Greece. Nick's parents have decided that in this new place, they want to keep Nick at home with them, and not have him be at the school with the other children. The family speaks only Greek, and they think Nick will do better if he sticks to his family's language and doesn't try to learn English. They say they can teach him everything he needs at home.
>
> What should happen? Is it okay for the Costa family to live in the town but keep Nick at home, or should they be required to send their son to the town school like all the other families do?

Phase 1 individual assessment. The positions of all participants on this topic were elicited via an individual written questionnaire. The questionnaire asked the participant to indicate their position among the three options, "Nick should go to the school," "Nick can be taught at home," and "undecided." Participants were then asked to indicate how certain they were of their positions on a 6-point scale ranging from *certain* to *not sure at all.* The questionnaire then proceeded to pose the same sequence of probing questions described earlier for the CP-topic individual assessment.

Phase 1 dialogic engagement. Based on responses in the individual assessment, two groups of 14 students each were formed, one group who chose the homeschool option and another group who either chose mandatory townschool ($n = 9$) or were undecided but noted reasons supporting the townschool side ($n = 5$). Within each of the groups, homeschool and townschool, same-gender pairs were formed. To the extent possible, an additional criterion used in forming these pairs was to pair participants who provided different reasons for the same position.

The same-side pairs thus formed remained together until the "showdown" segment (see below) of Phase 1. Beginning at the next class session, pairs in one group assembled in one classroom and pairs in the other group assembled in an adjacent classroom. Each pair was assigned to conduct a dialog via IM with a pair from the other room. Their goal, they were instructed, was to try to convince the other pair that their position was the better one. Each pair was further instructed to collaborate with their partner to decide what they wished to say to the opposing pair and, once they were in agreement, to type their response and send it to the opposing pair.

Those in one room were arbitrarily chosen to begin the dialogs by stating and supporting their position, with the pairs in the other room then given the opportunity to respond. Each of the pairs in one room thus continued their dialogs with one of the pairs in the other room, with dialogs lasting an average of 25 min. At the next session, each pair conducted a dialog with a different opposing pair, until each pair had debated every opposing pair—a total of seven sessions. One or two adult coaches circulated each room during these sessions, answering any questions and reminding pairs to collaborate with one another in deciding what response to make.

Participants received no explicit instruction with respect to argumentation. The only exception, used occasionally if it appeared in danger of being violated, was to "criticize ideas, not people." The only implicit instruction was the invoking of a metaphor expressed occasionally to the whole group that in responding they should strive to "get the ball over the net and back to the other side."

Phase 1 reflective activity. After four of the seven dialog sessions had been completed, the first reflective activity was introduced at the next session. At these sessions, each pair was given a printed transcript of the preceding session's dialog. (Because sessions took place twice a week, the interval between the dialogic session and the reflective session was 3–4 days.)

In the reflective session, in addition to the transcript of the preceding session, pairs were given one of two scaffold sheets. The "Other's Argument" scaffold sheet prompted pairs to examine the opposing pair's argument and their own counterargument. The "Own Argument" scaffold sheet prompted pairs to examine the opponent's counterargument and their own rebuttal to this counterargument. In each case, they were asked to contemplate what a better counterargument or rebuttal might have been. The adult coaches circulated during these sessions to provide clarifications if needed.

Reflective sessions took place a total of three times during Phase 1, interspersed among the dialog sessions. Specifically, the first Other's Argument session occurred after the fourth dialogic session, the first Own Argument session occurred after the fifth dialogic session, and a combined reflective session in which both scaffold sheets were available occurred after the sixth dialogic session.

Phase 1 preparation and showdown. Following the seven dialogic and three reflective sessions, participants were told at the next session that it was time to begin preparation for a final showdown debate with the opposing side. They were given one session to prepare for the showdown and the following session to conduct it.

To prepare, pairs within their opposing-sides classrooms were regrouped into an A and a B team. One team was encouraged to become "own argument" specialists by reviewing the possible counterarguments the opposition would make to their arguments and the rebuttals they would have ready. The other team was encouraged to become "other argument" specialists by reviewing the possible arguments the opposition would make and the counterarguments they would have ready. Transcripts of the dialogs, the previously completed scaffold sheets, and blank index cards were made available.

For the showdown, the two opposing groups remained in their respective classrooms and communicated with one another via the same IM software but this time with a single computer terminal and Smartboard in each room. To conduct the showdown, students were again reorganized, this time into Team A and Team B (each containing similar numbers of "own argument" and "other argument" specialists), with Team A allocated to be the acting team for the first half of the activity (with Team B advising by transmitting suggestion cards to the acting team), and the roles reversed for the second half of the activity. The entire activity lasted 50 min.

At the session following the showdown, an argument map that had been prepared by the researchers was presented to the assembled class. Different colors, as well as labels, were used to identify argumentive moves as effective (green), less effective or ineffective (red), and neutral (blue or black). A point system was applied to this map, making it possible to declare a winning team.

Phases 2 and 3. Phase 2 began following winter vacation and Phase 3 following spring vacation. Phases 2 and 3 proceeded in an identical way to Phase 1 except for new topics and a slight reduction in number of sessions (one less dialogic and one less reflection session) due to the slightly reduced time available in the school calendar for these two phases.

The topic for Phase 2 was whether students who continually disrupt the classroom should be expelled or allowed to remain. The topic for Phase 3 was whether teachers should receive experience-based pay (with more experienced teachers more highly paid) or all teachers should receive the same pay.

Final Assessment

The final assessment occurred following Phase 3. It was on the same (CP) topic and contained the same components as the initial assessment. In the dialogic activity, all participants argued with the same opponents they had engaged in the initial assessment.

Results

We begin analysis with a comparison of participants' performance on initial and final assessments. This analysis allows us to establish the overall progress in skills that each participant made, assessed in a content domain different from the ones involved in the intervention. In addition to its importance in its own right, this information proves useful in examining progress during the intervention itself. Our initial analytic task, however, before microgenetic examination could begin, was to apply a common dialogic coding scheme to the initial and final dialogs, as well as to the intervention dialogs.

Coding of Electronic Discourse

Each of the 161 electronic dialogs (7 per session at 19 intervention sessions across the three intervention topics and 14 each at the initial and final assessments) was analyzed based on an elaborated version of the scheme used in earlier research (see Felton, 2004; Felton & Kuhn, 2001; Kuhn & Udell, 2003; Udell, 2007). The discourse is segmented into idea units and a code assigned to each unit based on its functional relation to the opponent's immediately preceding utterance in the dialog. Coders blind to the identity of participants or time of the dialog took part in the coding. Interrater reliability (calculated on a randomly chosen 15% of the intervention dialogs and 100% of the initial and final dialogs, all of which were double coded) indices were 89% agreement on individual units within dialogs and a Cohen's kappa of .87. Coding reliability was examined individually for each of the major categories to be examined here and was found to be in a comparable range. The percent agreement for the Counter-C category (see definition below), for example, was 98% with a Cohen's kappa of .89. An additional coding scheme for meta-level discourse was developed to aid in examination of the intervention dialogs and will be described in presenting the intervention data analysis.

Skills at Initial and Final Assessments

At the initial assessment, dialogs contained a mean of 14.94 coded utterances ($SD = 8.26$). At the final assessment, dialogs contained a mean of 15.06 utterances ($SD = 8.12$), a nonsignificant difference. Mean number of words contained in each utterance increased, however, from 6.83 ($SD = 2.57$) at the initial assessment to 11.70 ($SD = 4.44$) at the final assessment—a significant increase, $F(1, 54) = 25.28$ ($p < .001$).

In comparing the kinds of utterances contained in dialogs at the two assessment times, only those categories from the coding scheme are examined that accounted for greater than 5% of utterances, averaged across dialogs, by the time of the final assessment. These categories are Clarify, Case?, Counter-A, and Counter-C. Percent use of each of the other categories was 5% or below at one or both assessment points. (Meta-level categories are addressed later.) In the comparisons across time reported below, an arcsine transformation was performed to normalize proportions.

Counterargument (Counter-A and Counter-C). Counterargument is the central and most essential skill to be mastered if one is to engage in effective argumentive discourse. Its use reflects awareness of the dual goals of argumentive discourse noted earlier (to weaken the opponent's argument, as well as to gain concessions from the opponent that will support one's own argument). We assess the use of two forms of counterargument. The weaker form, Counter-A (counter-alternative), addresses the other's position and undertakes to reduce its force, but it does so by introducing a new criticism of the other's position, not directly related to the argument the other has just made in support of his or her claim. The stronger, more skilled form of counterargument, Counter-C (counter-critique), by contrast, directly addresses the other's argument, criticizing it in a way that seeks to weaken its force. In the case of the CP topic, for example, if the other's argument is "CP will keep dangerous criminals off the street," a common Counter-C is "Life imprisonment can do that as well," whereas a common Counter-A is "Sometimes innocent people are executed."

In analyzing change in counterargument usage from initial to final dialogs, we examined three indices of counterargument skill. The first is the proportion of utterances a participant made that were counterarguments of either type. The second is the proportion of utterances that were coded as Counter-Cs. The third is the proportion of utterances coded as counterarguments that were Counter-Cs. Thus, we look both at mastery of counterargument itself and then at indications of the extent to which counterarguments are being used most effectively.

All three of these indices showed significant increases from initial to final assessments. The percentages were transformed using an arcsine transformation. Overall Counter usage increased from an average of 6.34% ($SD = 7.66$) of utterances at the initial assessment to 26.53% ($SD = 17.27$) at the final assessment, a significant change over time, $F(1, 27) = 31.75$, $p < .001$, $\eta_p^2 = .540$. Counter-C usage increased from 4.50% ($SD = 7.09$) to 21.37% ($SD = 14.61$), $F(1, 27) = 19.89$, $p < .001$, $\eta_p^2 = .424$. Finally, the percentage of Counters that were Counter-Cs increased from 31.25% ($SD = 45.45$) to 63.79% ($SD = 37.75$), $F(1, 27) = 7.27$, $p = .012$, $\eta_p^2 = .212$.

As important as these group trends are patterns of individual change, especially as we wish to relate such patterns to patterns of individual change that participants show during the intervention itself. At the initial assessment, only 36%—10 of 28—participants made any Counter-C arguments. At the final assessment, all but one did so (a significant change, McNemar test, $p < .001$). At the initial assessment, 50% (14 of 28) participants made *any* counterarguments at all, but all did so at the final assessment (also a significant change, $p < .001$).

Also of interest are the proportions of participants who showed increases in their overall Counter and Counter-C arguments from initial to final assessments—a majority did so, 85% in the case of overall Counters and 78% in the case of Counter-Cs. Only 1 participant showed no Counter-C usage at either assessment. Of the remaining participants who failed to show improvement in Counter-C usage, all showed some slight usage at each assessment.

Rebuttal. A further important aspect of counterargument is the extent to which discourse participants are able to maintain focus and consistency in intent and execution that enables them to sustain the critique of one another's arguments. Once the initiator of a claim responds to an opponent's counterargument by countering this counterargument, we have what is commonly referred to as a rebuttal. In our analyses of participants' dialogs, we looked at both frequency of rebuttals and length of rebuttal strings, that is, over how many contributions to the dialog a sequence of Counters is maintained.

Participants in the present study showed little initial skill in rebuttal. Only 6 of 28 (21%) made any rebuttal utterances at the initial assessment. At the final assessment, this number increased to 19 (68%)—a threefold and significant increase ($p < .001$). Consistent with this increase, mean number of rebuttals across participants increased significantly from 0.39 ($SD = 0.88$) at the initial assessment to 1.82 ($SD = 2.06$) at the final assessment, $F(1, 26) = 10.51$, $p = .003$. Overall frequencies are presented in this case, rather than percent usage, because production of rebuttals is affected by the opportunities available to make them. In other words, a participant cannot make a rebuttal if the opponent has made no counterargument. We therefore also calculated the percentage of rebuttal opportunities that were taken advantage of, that is, the number of rebuttals an individual made relative to the number of opportunities for rebuttal offered by the opponent. This percentage increased from an average of 17.65% ($SD = 35.15$) at the initial assessment to 43.79% ($SD = 36.48$) at the final assessment. This change, note, however, is confined to a reduced sample, the 12 participants who had any opportunities for rebuttal at the initial assessment (all did so at the final assessment), and this sample size is not large enough for statistical test. Interpretation of group data on increasing rebuttal skill thus remains limited until counterargument skill reaches a level at which all participants have available numerous opportunities to make rebuttals.

We looked finally at the length of rebuttal strings, whenever rebuttals were made. A length of 1 represents a sequence consisting of assertion – counterargument – counterargument; a length of 2 represents a sequence consisting of assertion – counterargument – counterargument – counterargument. Mean length rose only slightly from 1.83 at the initial assessment to 2.13 at the final assessment (a statistically nonsignificant increase). These numbers show a typical successful effort on the part of a maker of an initial counterargument to not allow to stand as successful the opponent's move to weaken this counterargument (length of 2, i.e., assertion – counter – counter – counter).

Exposition (Clarify and Case?). Clarify, the most common utterance category, includes all those utterances in which the participant asserts or elaborates his or her own arguments. Although in previous work (Kuhn & Udell, 2003) usage in this category has declined as counterargument usage increased, participants in the present study overall showed a constant percent usage of Clarify—an average of 21.6% ($SD = 10.1\%$) at the initial assessment and 21.7% ($SD = 14.2\%$) at the final assessment. Because overall length of dialogs (reflected in number of utterances) remained constant, as reported earlier, this finding indicates absence of change in either number or proportion of utterances in the Clarify category.

Case? is another category that in past research has shown different patterns of change in different samples, with the suggestion that its use may increase temporarily but then diminish as other argument strategies become available (Felton, 2004; Felton & Kuhn, 2001). In Case?, the participant poses a situation to the other and asks the other to respond. In discussing the CP topic, for example, one member of the dyad might ask the other how she would feel about CP if the murder victim were a family member. The use of Case? is effective when it poses the situation with the intent of securing concessions from the other and in this respect may be seen as a stepping stone or alternative to direct counterargument. In the present analysis, the average percent use of Case? increased from 1.54% ($SD = 3.10$) at the initial assessment to 6.84% ($SD = 11.68$) at the final assessment, $F(1, 27) = 13.18$, $p = .001$, $\eta_p^2 = .328$.

Individual Argument

Although our purpose in the present work is to examine the development of dialogic argument skills in their own right, we did obtain evidence that increasing dialogic skill leads to improvements in individual (nondialogic) argument. Such evidence corroborates our own and others' earlier studies of individual argument skills (Kuhn, Shaw, & Felton, 1997; Reznitskaya et al., 2001) in which such positive effects of practice in dialogic argument on individual argument have been observed (allegedly due to enhanced awareness of the "missing interlocutor" implicit in nondialogic arguments).

The coding scheme developed by Kuhn et al. (1997) was used for assessment of individual CP arguments. In particular, we examined the quality, quantity, and sidedness (two- vs. one-sided) of initial and final individual arguments for or against CP. As 1 participant was away for medical reasons at the final individual assessment, these analyses are based on a sample size of 27. Level 1 reasons in this scheme offer genuine justifications for or against the use of CP, whereas Level 2 reasons offer only functional arguments regarding the criteria for use or nonuse of CP (e.g., "only if it's a serious crime"), without addressing justification of the practice itself, and Level 3 reasons are entirely nonjustificatory (e.g., appeal to precedent or authority). (See Kuhn et al., 1997, for the full scheme.)

At the initial assessment, a mean of 0.56 (*SD* = 0.93) different Level 1 reasons were produced compared to a mean of 1.33 (*SD* = 1.00) at the final assessment—a significant increase, *t*(26) = 2.95, *p* = .007. (Repetitions of the same argument code did not increase an individual's score.) Of the 27 participants, 25 added at least one new Level 1 reason at the final assessment. Total number of different reasons (of all levels) increased from a mean of 3.07 (*SD* = 1.59) to a mean of 4.22 (*SD* = 1.80)—a significant increase, *t*(26) = 2.37, *p* = .026. The number of participants producing two-sided arguments (citing reasons both for and against CP), either initially or when elicited ("Can you imagine someone having another position that is different from your own? What would it be?") increased from 17 to 22 across the two assessments.

Dialogic Skill Development During the Intervention

Analysis of the discourse patterns during the three phases of the intervention is based on a sample size of 28. We divide the presentation of results into two sections, the first section reporting on trends in discourse performance itself and the second section examining trends in the use of meta-level discourse that seeks to evaluate, question, monitor, direct, or otherwise manage this performance.

Overall patterns of change at the group level. For these analyses, the dialog sessions for each topic were collapsed into beginning, middle, and final segments (with two sessions combined in each category except for the middle category for Topic 1, which included three sessions). A participant received the same score as his or her (same-side) partner for a particular session (and the pair taken as the unit of analysis where appropriate in statistical analyses). Dialogic sessions remained fairly stable over time with respect to length. Mean number of utterances per session were 42.12 (*SD* = 8.22) for Topic 1, 37.36 (*SD* = 9.21) for Topic 2, and 42.71 (*SD* = 10.80) for Topic 3. Mean length of utterances increased only slightly from 8.11 (*SD* = 6.17) for Topic 1 to 9.55 (*SD* = 8.10) for Topic 2 and 9.17 (8.52) for Topic 3. Mean word length is overall higher, however, for the Counter-C category, a major focus of our interest, and increased across topics, from 11.29 (*SD* = 6.85) for Topic 1 to 13.75 (*SD* = 8.61) for Topic 2 to 15.74 (*SD* = 9.45) for Topic 3.

Counterargument. In Table 1, there appear the mean percentages of all utterances that were coded as Counter (top row) or Counter-C (second row). The final row shows the mean percentages of Counters that were coded as Counter-C. As reflected in the leftmost column (beginning of Topic 1), the mean percentage of utterances that a participant devotes to counterargument increases substantially, to 24.43%, from the earlier reported mean level of 6.34% at the initial (CP) assessment, where participants argued alone against their opponent rather than with a partner as they do here. Similarly, the mean percentage of Counter-Cs increased from the initial CP assessment, from the earlier reported 4.50% when working alone to the 13.87% shown here. These changes are suggestive of the advantages provided by collaboration with a partner as well as what we have observed in past work to be the motivating and focusing effects of working toward the social objective of the showdown.

As seen in Table 1, however, performance continues to improve during Topic 1 but does not increase further in these respects after the first topic. Based on pairs as the unit of analysis, for Topic 1, gains (from beginning to end) are significant for overall Counters, Z = −2.605 (Wilcoxon test), *p* = .009, effect size *r* = −.696, and for Counter-Cs, Z = −2.166, *p* = .03, effect size *r* = −.579. For Topic 2, gains are significant only for Counter-Cs, Z = −2.355, *p* = .019, effect size *r* = −.629 (for overall Counters, Z = −1.789, *p* = .074), and for Topic 3, no trends over time are significant (Counter-C: Z = −0.722, *p* = .470; overall Counters: Z = −0.659, *p* = .510). Note finally that when we

Table 1
Mean Percent Usage of Counterargument and Counterargument-Critique (Counter-C) Across the Three Topics

	Topic 1		Topic 2		Topic 3	
	Beginning	End	Beginning	End	Beginning	End
Total Counters	24.43 (8.50)	32.30 (10.49)	16.19 (5.80)	20.77 (8.55)	20.15 (8.98)	18.70 (8.17)
Counter-C	13.87 (4.89)	20.62 (9.49)	9.28 (4.99)	14.31 (8.17)	10.37 (5.74)	9.32 (7.75)
Percentage of Counters that are Counter-C	58.22 (15.26)	63.30 (17.07)	57.29 (29.52)	61.16 (25.28)	50.64 (27.35)	48.23 (29.02)

Note. Standard deviations appear in parentheses.

compare the percentages in Table 1 to those for the final CP assessment reported earlier (26.53% for Counters and 21.37% for Counter-Cs), it is apparent that gains made during the intervention are maintained when participants return to dialogic arguments against an opponent without the aid of a collaborative partner.

Rebuttal. We do not undertake statistical analysis of changes in percentage of rebuttal opportunities successfully met due to the initial scarcity and only gradual increase in the availability of such opportunities over time, as noted earlier in the report of rebuttal usage at initial and final (CP) assessments. By the time of the third intervention topic, however, rebuttal skill had been attained to the extent that all pairs showed some rebuttal usage during this topic and 41.52% of rebuttal opportunities were successfully taken advantage of (with all participants having such opportunities)—a percentage comparable to the 43.79% noted earlier observed at the final (CP) assessment. The mean length of Topic 3 rebuttal chains was 2.03 (range = 1–7) comparable to the 2.13 (range = 1–5) observed at the final (CP) assessment. An example of a successful rebuttal chain (of Length 2) is presented in Table 2 for Topic 3 (teacher pay).

Individual change patterns. Although the preceding results are suggestive, they do not tell us to what extent individual patterns of performance conform to group trends portrayed in Table 1. For this purpose, we looked at individual change patterns. We focus on the Counter-C indicator, as our most important index of growing skill in argumentive discourse. For purposes of examination, we also divide participants into two groups, the 21 who had shown progress from initial to final CP assessment (based on Counter-C usage) and the 7 who had not.

In Figures 1a and 1b, we present two examples of individual change patterns among the majority group of 21 who showed increase in Counter-C usage from initial to final assessment. In one case (Participant 25), the individual shows no Counter-C ability at the initial assessment but then, in collaboration with successive partners, exhibits some Counter-C usage in engaging with the three topics; at the final (CP) assessment, again working alone, she is now able to produce some Counter-C without a partner's collaboration.

Table 2
Example of Rebuttal Chain for Topic 3 (Teacher Pay)

Pair A: WE THINK THAT THEY SHOULD GET THE SAME PAY, BECAUSE SOME TEACHERS MAY LEAVE (*coded Clarify*)

Pair B: but the teachers won't leave if they think they get more pay next year + It will make them want to stick to the job (*coded Counter-c*)

Pair A: THE TEACHER'S WITH LESS EXPERIENCE WON'T THINK IT'S FAIR. BECAUSE NEXT YEAR IS A WHILE AWAY (*coded Rebuttal*)

Pair B: then they can attract experienced teachers (*coded Rebuttal*)

Another participant (Participant 10, Figure 1b) does show some Counter-C usage when working alone at the initial assessment, continues to display it to a variable degree during the intervention, and then increases this usage near the end of the intervention and in transferring back to the solo context. In both cases, then, the social collaboration experience appears to have been fruitful, and the level of performance does not revert to its original level when the social context is removed. These cases illustrate patterns most common among the groups of 21 who progressed from initial to final CP assessment.

Among the 7 participants who did not show any indication of progress from initial to final CP assessment, 4 showed equivalent performance from initial to final assessment and 3 declined slightly, with performance remaining at a low level in all cases (and 0 in one case, shown here). In all these cases, more such usage appeared during the intervention while the participant was working collaboratively, but the participant did not transfer it to the solo context of the final CP assessment. One of these cases appears in Figure 1c (Participant 28).

Within the three topics, patterns of individual performance resemble those of the three illustrative participants in Figures 1a to 1c. Variability is the norm, rather than consistent upward change, within each topic, and performance achieves its maximum levels during the first topic (as also reflected in Table 1)—an unanticipated finding we return to in the discussion of our findings. Despite this variability, there appears within the intervention sessions, at least for the 21 "progressing" participants (the group defined by progressive change across CP assessments), a common trend of progress within a topic from the beginning to final sessions of engagement with that topic. To quantify this trend, we defined a pattern as upwardly progressing over the beginning–middle–end phases of engagement with a topic if performance during the second was higher than the first and performance on the third higher than the second. We defined a pattern as mixed upward if performance at the final phase was higher than at the beginning phase, but the two intermediate links (to the middle point) reflected either plateau or decline.

Using these definitions, among the 21 progressing participants (the group defined by progressive

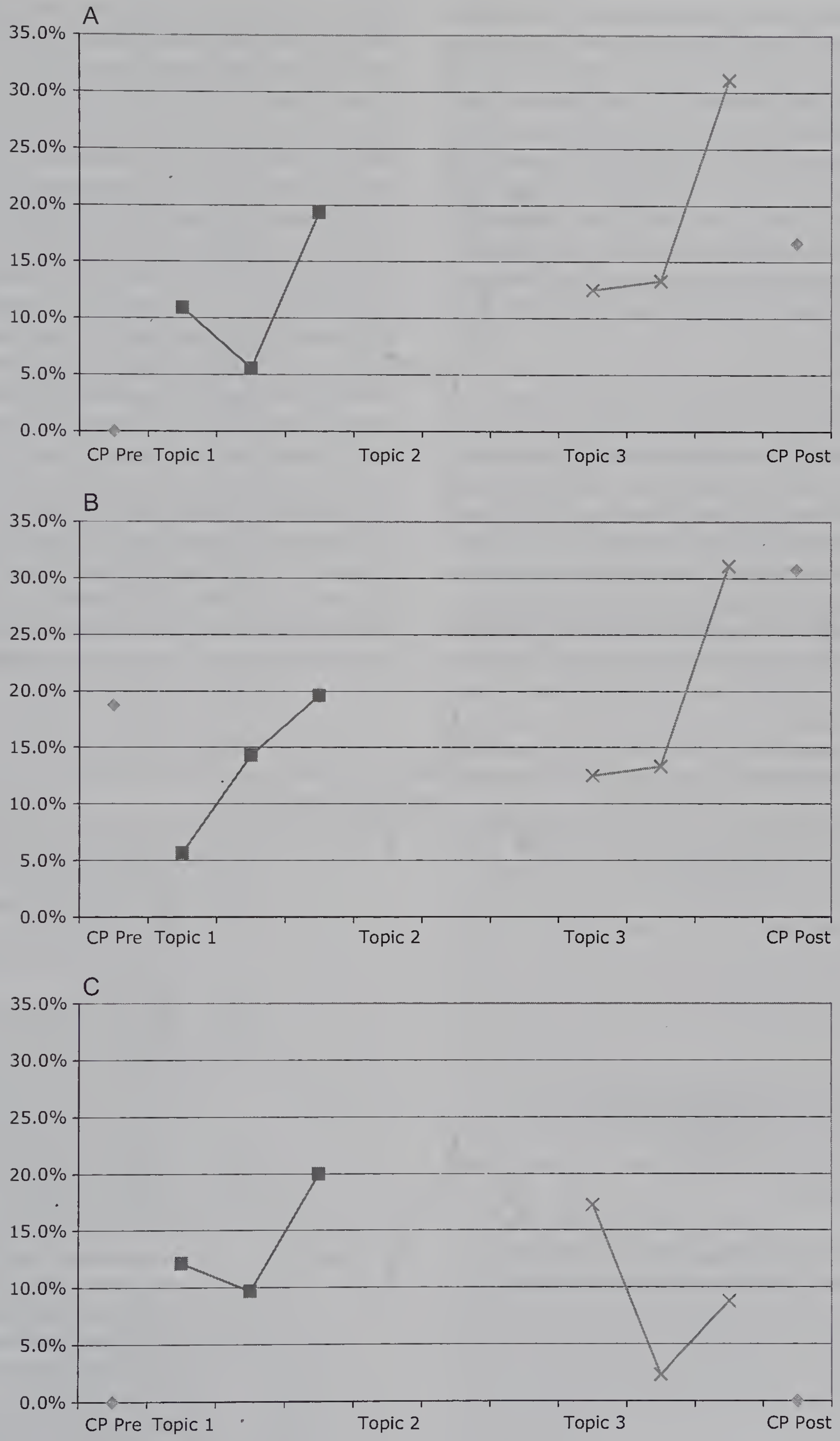


Figure 1. Percent Counter-Critique usage for discourse use across topics: (a) Participant 25, (b) Participant 10, and (c) Participant 28. CP = capital punishment.

change across the two CP assessments), the majority of patterns were in one of these two categories (upward or mixed upward)—specifically 73%. In contrast, among the remaining 7 participants who did not show progress across the two CP assessments, this percentage was 48% (roughly what would be expected if change were in no consistent direction). Although sample size is too small for meaningful statistical analysis, these findings are of interest in suggesting a link between patterns of performance during the intervention and at the initial and final (CP) assessments.

Meta-Level Development During the Intervention

Consistent with the questions we initially identified as central to the present work, our objective in this phase of the analysis was to examine the extent to which participants showed progress in recognizing the goals of argument, as distinguished from progress in implementing successful argumentive strategies. We expected progress of the former type to be most visible when participants collaborated with a partner toward a collective goal of defeating the opposing side in the showdown activity (devised to highlight this goal). We did not have resources in the present study to record the conversation between collaborating partners during the dialogs as they formulated the responses they would make to their opponents. These responses themselves, however, became a matter of record and could be examined with respect to what evidence they showed of meta-level understanding of argumentive goals and strategies. Meta-level statements we defined as statements *about* the dialog in contrast to statements that constitute the dialog. Omitted from this analysis are utterances that were meta-level but procedural in nature, for example, requesting or receiving clarification regarding task instructions or software issues and did not pertain to the dialog itself.

To further examine the nature of these meta-level statements, we developed the coding scheme shown in Table 3. The subcategories of meta-level statements appearing in Table 3 are listed in descending order with respect to their centrality to understanding of the goals and purposes of argument together with their

Table 3
Types of Meta-Level Statements

Type	Definition	Overall percent usage	Examples
Meta-directive	About the dialog, with a directive to opponent	7.59	—please state your argument —give us some reasons —give us proof and reason —Then prove it . . . —you need to give facts instead of opinion —STAte why —LETS GET BACK ON TOPIC —Please answer my question —please answer 2 our arguments!
Meta-argumentation	About the dialog, without an explicit directive to opponent	2.99	—you have not answered —u havent given good answers + so we r gonna keep askin it —that's opinion not a fact —how do you know —We're not talking about that
Meta-argument	About respective positions rather than the argumentation process	4.85	—thats our point —thats what WERE arguing for! —that is our main argument —Yeah thats what weve been trying to say —its the same argument except with some modifications —look we agree on something! —you guys agreed with us before —we never said that
Meta-comprehension	About comprehension	0.82	—You are not making sense —we don't understand —I don't get what you say man

Table 4
Mean Percent Usage for Total Meta and Meta-Directive Across Topics

	Topic 1			Topic 2			Topic 3		
	Beginning	Middle	End	Beginning	Middle	End	Beginning	Middle	End
Progressing participants									
Meta-directive	4.62 (3.90)	5.84 (5.36)	4.91 (3.64)	9.51 (6.49)	14.84 (9.59)	8.43 (7.69)	11.20 (6.28)	10.70 (9.16)	7.55 (4.77)
Total meta	15.27 (5.53)	13.39 (8.92)	11.48 (5.67)	16.64 (7.56)	25.00 (12.48)	16.84 (10.18)	25.18 (12.29)	24.21 (12.37)	21.57 (7.24)
Nonprogressing participants									
Meta-directive	6.03 (3.20)	4.82 (2.45)	6.12 (4.99)	10.15 (5.30)	6.56 (7.72)	7.43 (6.09)	10.27 (7.23)	18.09 (13.39)	9.06 (4.62)
Total meta	13.50 (5.09)	13.05 (9.37)	12.12 (9.38)	15.77 (7.17)	12.17 (5.31)	14.93 (8.53)	19.31 (8.38)	24.85 (15.72)	22.00 (10.56)

Note. Standard deviations appear in parentheses.

overall frequencies of usage. After using a portion of the data to develop the scheme, two of the authors independently coded one third of the data, achieving a percent agreement of 82% for identification and categorization of meta-level statements, with a Cohen's kappa of .69. Differences were resolved by discussion, and all remaining statements were coded by one of these two coders.

Overall change at the group level. To conduct an analysis of change over time in the various kinds of meta-level statements, as well as overall usage of meta-level statements, we again examined separately the participants in two groups, the majority group ($n = 21$) who had shown positive change in the initial to final CP assessment, and the minority group ($n = 7$) who showed no indication of such change ("progressing" and "nonprogressing" group, respectively). In Table 4, there appears for each group mean percentages of usage across time of utterances of a meta-level type. (The time intervals shown are the same as those in the earlier reported analyses in Table 1.) As seen in Table 4, for the majority (progressing) group, the mean percentage of meta-level statements increased from 15.27% at the beginning of Topic 1 to 21.57% at the end of Topic 3, a significant increase, with gains for the minority group comparable. These gains are significant for the sample as a whole, $Z = -3.552$ (Wilcoxon test), $p < .001$, effect size $r = -.671$, as well as for the progressing group examined individually. (Note, however, that in contrast to the analysis of Table 1 data, the individual, rather than the pair, must be the unit of analysis, as pair composition changes across topics. Given the lack of independence between members of a pair, significance levels should therefore be interpreted conservatively.)

Change over time was also examined for each of the subcategories in Table 3. Change is significant only for the first of the subcategories—Meta-directive. For the sample as a whole, $Z = -2.619$ (Wilcoxon test), $p = .009$, and effect size $r = -.495$ (although the same caution noted earlier applies).

These gains, however, were not maintained when participants returned to arguing against an opponent without a partner in the final (CP) assessment. Meta-level coding of the CP dialogs showed a steady rate of meta-level usage at the two points, for example, for meta-directives, 5.6% at initial and 5.8% at final CP assessment.

Individual change patterns. Examination of individual patterns of performance with respect to meta-level usage confirms that the pattern of increasing meta-level usage, in particular the Meta-directive category, is consistent across participants. Only 2 of

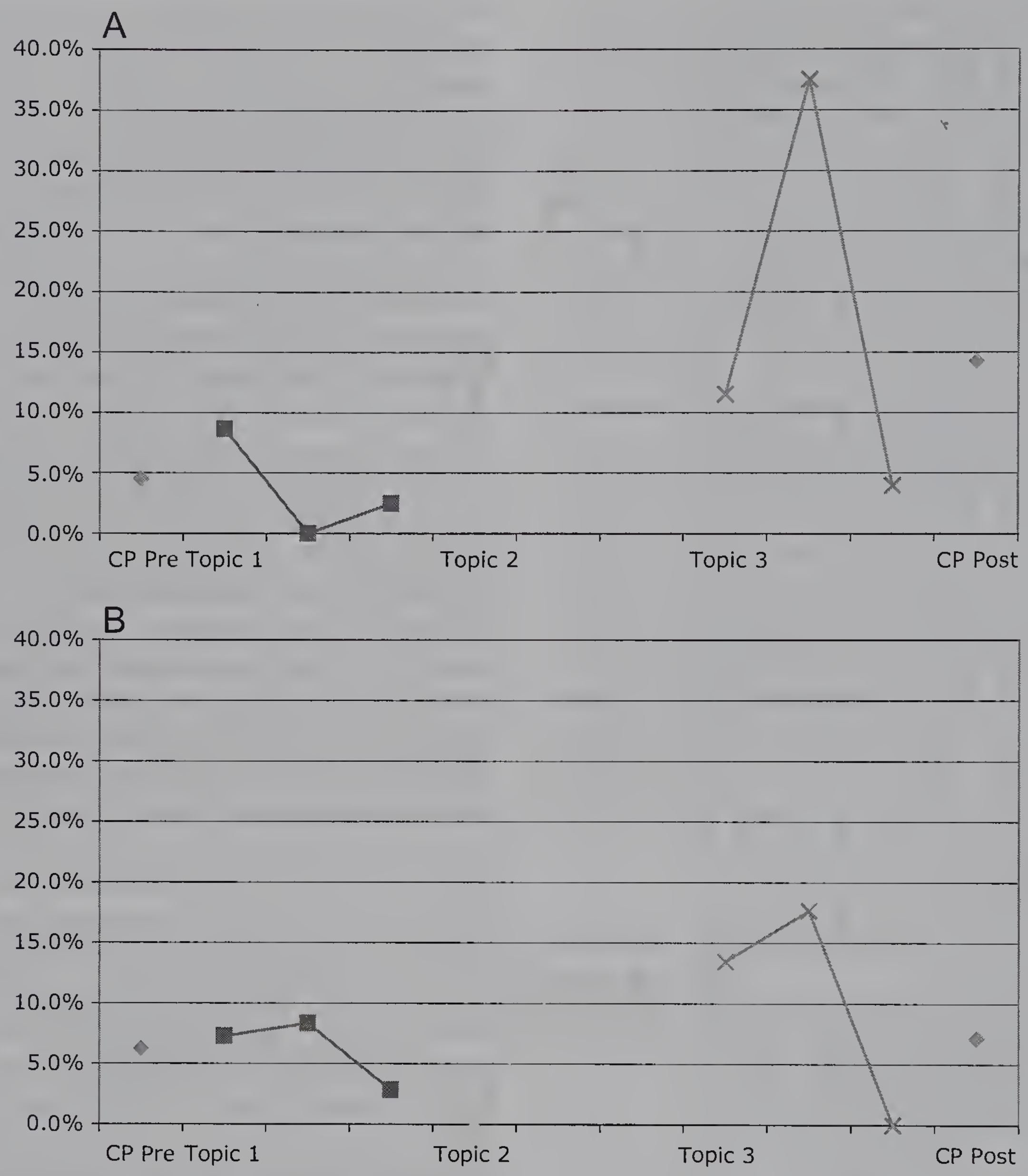


Figure 2. Percent meta-directive usage for meta-level discourse across topics: (a) Participant 19 and (b) Participant 11.

28 participants in fact fail to show such a pattern. Two typical examples appear in Figures 2a and 2b (Participant 19) and (Participant 11). As seen there, the progress reflected in Figure 2b is less dramatic than that in Figure 2a. In both cases, however, there appears the typical pattern of a decline in Meta-directive usage when the participant no longer works with a partner in the final (CP) assessment.

Implications of meta-level change. In summary, almost all participants showed increasing meta-level usage during the course of their collaborative argumentation activities. At the final dialogic assessment, no longer working collaboratively, three quarters of participants (21 of 28) showed enhanced argumentation skill (indexed by percent use of Counter-C). Did this increase in meta-level discourse about argumentation reflect understanding that contributed to the enhanced skill observed in the conduct of argumentation? Our theoretical assumptions regarding the importance of meta-level regulation of behavior suggest a positive answer, but it is not straightforward to identify strict causal evidence of such an effect. Because meta-level discourse emerged and increased over time in almost all participants, we cannot compare those who did and did not engage in meta-level discourse. Furthermore, participants worked in pairs during the intervention, and we cannot be sure that the meta-level discourse directed to the opponents was contributed to equally (or at all) by each member of the pair (and would therefore be expected to affect the individual skill assessment of both members of the pair).

In search of some further evidence of the contributory role of meta-level processes, we examined pairs who showed more versus less than the typical proportion of meta-level discourse to assess whether these two groups would differ in argumentation skill level. For this analysis, we chose Topic 2, when meta-level discourse was most likely to have emerged, after participants had become familiar enough with the activity to begin to reflect on how best to carry it out. We divided the Topic 2 pairs (partners differed, recall, for the other two topics) into those who showed relatively high proportions of statements in the meta-directive category (15% or more) and those who showed lower proportion of statements in the meta-directive category (less than 15%). Of the 14 pairs, 8 fell into the high category and 6 into the low category. For each pair, we calculated the mean of the 2 pair members' proportion usage of Counter-C at the final (CP) dialogic assessment. Among the 8 high meta-directive pairs, 7 showed Counter-C usage equal to or greater than 15%. Among the 6 low meta-directive pairs, only 1 pair showed Counter-C usage equal to or greater than 15%—a significant difference (Fisher exact test, $p = .026$). Although this comparison involves small numbers and should only be regarded as suggestive, it does offer some evidence of a link between meta-level discourse and argumentation skill.

Showdown

Although we did not formally analyze performance in the showdown preparation sessions, we did undertake the same discourse analysis of the three showdown transcripts as was performed on the intervention dialogs. Here, we saw that the focused group effort produced a higher level of skill than that observed during the dialogs between pairs. The overall percentages of showdown Counter-C statements were 58% for Topic 1, 61% for Topic 2, and 70% for Topic 3. To illustrate the level of performance that participants achieved as a group, we present in the Appendix the (slightly edited) showdown debate produced for the homeschool topic in the form of the argument map presented to students at the session following the showdown for this topic. The Appendix contains a black and white version (with effective moves in bold and ineffective in dotted lines) rather than the colored version (with green and red boxes) that students saw.

Discussion

Our findings highlight the value of the microgenetic method in providing a clearer and fuller picture of the changes that occurred among our participants than would have been evident from the pretest–posttest assessments alone. We therefore discuss our results in the reverse order in which they were presented, beginning with the microgenetic data on meta-level change and then proceeding to change at the strategic level and finally to the pre- and posttest assessments of first dialogic and then individual argument skills.

On virtually all the dimensions we examined, some participants were identified who failed to show progress. This group was smallest in number, however, on the dimension of meta-level discourse. Of the 28 participants, 26 showed a pattern that involved increasing usage of meta-level discourse, in particular meta-directives to the opposing pair. As elaborated earlier, we see this meta-level discourse as significant in indicating at least implicit understanding of the functions and objectives of argument. When participants say to their interlocutors, "Give us some reasons" or "You have not answered," they are displaying implicit understanding of what contributors to a dialog need to do in order to engage effectively in argumentation.

It is notable that these increases on the part of almost all participants were evident only when participants were engaging in the dialogic task with a collaborating partner. When participants again worked without a partner (at the final CP dialogic assessment), these increases were no longer evident (as reflected in comparisons of initial and final CP dialogs). In the collaborative context, participants had the cognitive and social support of one another in formulating their contributions to the dialog and indeed needed to collaborate in order to perform their task. They also shared the social goal of prevailing over the opposing side in the showdown. Neither of these conditions existed when they argued individually with an opponent on a different topic, one that they had discussed only once before and without a particular objective. We cannot say with certainty whether under the latter conditions participants lost sight of argumentation objectives that they were aware of in the collaborative condition. We see as a more likely interpretation, however, that they simply perceived no need to express them in meta-level talk in the solitary condition. They had a task to do (to debate the CP topic with an opponent), and there was no one they needed to consult with as to how best to carry out the task, so they simply got on and applied their skills to the task, with little comment on their own or their opponent's performance. If so, our findings point to the importance of social context and goal structure in influencing performance—a conclusion also reached by Nussbaum (2005) in a study of argumentation in a college-student population. The association we observed between extent of meta-directive discourse and later individual argument skill nonetheless supports the view that this meta-level discourse was productive with respect to skill development.

Turning to the strategic skills themselves, it is notable that, in contrast to what was the case for the meta-level discourse just discussed, positive change over time did not diminish when participants return to the nonsocial context of working without a collaborating partner and without a social goal. Counterargument and rebuttal, which lie at the heart of argumentation, increased in frequency among three quarters of the participants from initial to final CP assessment, and all participants showed some skilled counterargument (Counter-C) at the final assessment (compared to only a third at the initial assessment). The increase in frequency of skilled counterargument (Counter-C), to 21.37% of discourse, is consistent with our earlier studies involving similar methods (Felton, 2004; Kuhn & Udell, 2003; Udell, 2007), and the amenability of argumentation skill to increased usage with exercise and scaffolding of various kinds is also consistent with other lines of research involving both children (Anderson et al., 2001; Nussbaum, 2003; Reznitskaya et al., 2001) and young adults (Clark & Sampson, 2005; Glassner & Schwarz, 2005; Nussbaum, 2005).

An unexpected finding, revealed by microgenetic analysis, is that strategic-level (although not meta-level) gains were concentrated in the first third of the intervention activity (Table 1). We cannot, of course, be certain that the gains documented at the final (CP) assessment would have been the same had the participants not had the further experience of engaging in the second and third phases of the intervention. Still, we are left with the question of why they did not show further strategic gain during the second and third phases.

Although numerous explanations are plausible (e.g., motivational factors or ceiling effects), the explanation we favor derives from our going back to the dialog transcripts and comparing their content across the three intervention topics. Although we did not intend it, there may have been a subtle difference across topics. The first topic, the homeschool topic, was characterized by well-defined opposing positions (the protagonist in the dilemma is either homeschooled by his parents or attends the townschool), either of which could plausibly lead to a satisfactory outcome. The second topic, school expulsion, also had well-defined opposing positions (the misbehaving students are removed from the school or are not). In this case, however, neither alternative can be expected to lead to a satisfactory outcome for the misbehaving students. In the words of a number of our participants on both sides of the issue, "That won't solve the problem." Participants on one side noted that these students would continue to disrupt if not removed, whereas participants on the other side often noted that the students would just cause the same disruption if removed and sent elsewhere to school. The latter group thus devoted considerable dialog to proposing other solutions (e.g., bringing in a social worker or counselor to work with the misbehaving students) that would resolve the problem.

What we may be seeing here, we believe, however, is a shift in goal structure, driven by the topic content, from the classic *argument* goal structure (having one's position prevail over that of an opponent) to the *dilemma resolution* goal structure of resolving the underlying problem that the two opposing sides address. Our dialogic coding scheme, which classifies the function, not the content, of utterances does not provide direct evidence of such a shift across topics, and it must remain speculative. Our discourse analysis of the three showdowns, however, showed that even if participants were diverted from counterargument/rebuttal in their paired dialogs during the later topics, their skill in counterargument was not compromised when they proceeded to the focused context of the showdowns.

The appearance of a shift in focus during the paired dialogs becomes even more pronounced in examining dialogs on the third topic, teacher pay. A subgroup of participants on the experience-based-pay side of the issue during the course of the dialogs identified as their position a third alternative, merit-based-pay, acknowledging weaknesses of the original experience-based-pay position. Some of their dialog was thus diverted to identifying and elaborating this new position and away from counterargument against the opponents' position.

Should we regard such shifts as advances, as refusals to accept the opposing positions as defined and the seeking of a "third way?" Such an interpretation is consistent with the ideas of philosophers such as Gilbert (1997), who defines a model of what he calls "coalescent argumentation," the goal of which is to enrich group understanding rather than pit sides against one another—the sort of model that has been proposed by some psychologists as a possibly more fruitful model for instructional interventions (Anderson et al., 2001; Nussbaum, 2005).

While appreciating Gilbert's model, as developmentalists, we are inclined to see the value of older children and adolescents gaining mastery of traditional argumentation structure, as a foundation for acquiring skill in more complex argument forms. In previous research (Kuhn & Udell, 2007), we have observed young adolescents rely on what we called an avoidant pattern in addressing very simple arguments, stating, for example, "I'd just play both," in response to a question about the merits of choosing soccer versus basketball as one's sport. Mastery of the traditional argument – counterargument – rebuttal structure involved in debating two opposing positions, in our view, is a component of, and hence a necessary foundation for, engaging in more sophisticated forms of "coalescent" argumentation. In its absence, inexperienced arguers may find themselves uncertain as to whether a potential contribution to discourse that they may be contemplating is simply an unanchored addition to the current talk about the topic ("Here's another thing that might be relevant") or has an identifiable function to perform in the discourse.

We note, finally, that advances in dialogic argumentation skill did show evidence of influence on the individual written arguments collected as part of the final assessment. Although demonstrating such effects was not the focus of the present work, it is of course a result one would hope for, as well as one that has been reported in previous studies of the effects of discourse on individual argument skills and/or conceptual understanding in both children younger than those in the present study (Anderson et al., 2001) and in adults (Nussbaum, 2005), as well as in our own earlier work with the age group studied here (Kuhn et al., 1997). Note that new, more adequate reasons in support of his or her position on CP are not ones a participant would have had an opportunity to be exposed to from an external source. The most likely interpretation of their appearance, in our view, is that it is the cognitive effort of considering multiple views and generating counterarguments and rebuttals with respect to them, as well as the benefit noted earlier of increased awareness of the "missing interlocutor," that is responsible for the shifts observed toward more adequate, higher level reasons to justify one's position and away from less adequate, lower level ones.

Because our major interest in the present work was on the development of dialogic argument skill, we included as the main focus of our initial and final assessments the dialogic argument skill that a participant displayed in the absence of the social support that characterized the intervention. As a result, participants had some, albeit very limited, dialogic experience with the transfer (CP) topic, and this experience could have contributed to the improvements observed in their ability to produce an individual argument in support of their position on the topic. Evidence against such an effect, however, comes from a study by Udell (2007), with the CP topic among a slightly older adolescent group, in which no improvements occurred in either dialogic or individual skills as the result of simply an initial assessment of these same skills several months earlier. In future work, a fuller assessment of individual argument skills, and how they are affected by advances in dialogic and meta-level argumentation skills, is warranted to further elucidate the suggestive gains observed in this study.

In conclusion, the microgenetic method employed here has shown that argumentation skills are amenable to development with engagement and practice in a scaffolded context yet without direct instruction. In light of the earlier noted weaknesses in argument skills reported by educators among students at the high school and even college level, this finding alone is notable, especially given that our results also support the idea of dialogic argument as a developmental pathway to individual argument skills of the sort emphasized in academic contexts.

In addition, however, the present work we believe points to the importance of meta-level awareness and understanding of argumentation rather than simply the implementation of argument strategies at the performance level. Understanding and appreciating the value of argument is an epistemological achievement (Kuhn & Park, 2005). In the broad sense, this means seeing the point of argument and coming to appreciate it as the ultimately most powerful means of influencing others' thinking. Only in so doing will one be disposed to invest significant effort in developing the skills that argument entails. Children need to learn that argument is more than something to be avoided. This understanding is not intuitively given. Certain cultural values (to "live and let live" and to show tolerance toward others' views and avoid risk of offending) work against it, and only a more advanced level of epistemological understanding supports it (Hofer & Pintrich, 1997, 2002; Kuhn, 2005; Moshman, 2005).

In the narrower sense, this meta-level understanding entails understanding of the objectives of argument and specifically of the relevance of the other person's position. If the opponent's position is not relevant, the process through which one achieves victory over the opponent cannot be regarded as one of argument. Kuhn and Udell's (2007) study of children's and adults' arguments found, in both groups (although to a greater extent among children) and in multiple (open- and closed-ended) formats, a preference for articulating one's own position over examining and addressing the other's position. As noted earlier, this finding supports the view that the developmental challenge may lie more in understanding the relevance of argument strategy than in executing it. If so, the present work suggests that such understanding is amenable to development.

Leitao (2003) showed college students a series of short essays together with additional sentences of counterargument and rebuttal to the main argument, which could potentially be inserted into the essay.

Students were divided in their opinions as to whether adding such counterarguments and rebuttals would improve the persuasiveness of an essay. Such findings highlight the importance of meta-level development with respect to argumentation. Although tracking this meta-level development is less straightforward than tracking development at the strategic level of performance, both seem essential.

References

Anderson, R., Nguyen-Jahiel, K., McNurlen, B., Archodidou, A., Kim, S., & Reznitskaya, A., et al. (2001). The snowball phenomenon: Spread of ways of talking and ways of thinking across groups of children. *Cognition and Instruction, 19*, 1–46.

Andriessen, J. (2006). Collaboration in computer conferencing. In A. O'Donnell, C. Hmelo-Silver, & G. Erkens (Eds.), *Collaborative learning, reasoning, and technology.* (pp. 197–231). Mahwah, NJ: Erlbaum.

Andriessen, J., Baker, M., & Suthers, D. (Eds.). (2004). *Arguing to learn: Confronting cognitions in computer-supported collaborative learning*. New York: Springer.

Bell, P., & Linn, M. (2000). Scientific arguments as learning artifacts: Designing for learning from the Web with KIE. *International Journal of Science Education, 22*, 797–817.

Brem, S., & Rips, L. (2000). Explanation and evidence in informal argument. *Cognitive Science, 24*, 573–604.

Chinn, C. A. (2006). Learning to argue. In A. O'Donnell, C. Hmelo-Silver, & G. Erkens (Eds.), *Collaborative learning, reasoning, and technology.* (pp. 355–383). Mahwah, NJ: Erlbaum.

Clark, V., & Sampson, D. (2005, June). *Analyzing the quality of argumentation supported by personally-seeded discussions*. Paper presented at the Computer-supported collaborative learning conference, Taipei, Taiwan.

Damon, W. (1984). Peer education: The untapped potential. *Journal of Applied Developmental Psychology, 5*, 331–343.

Felton, M. (2004). The development of discourse strategies in adolescent argumentation. *Cognitive Development, 19*, 35–52.

Felton, M., & Kuhn, D. (2001). The development of argumentive discourse skills. *Discourse Processes, 32*, 135–153.

Gilbert, M. (1997). *Coalescent argumentation*. Mahwah, NJ: Erlbaum.

Glassner, A., & Schwarz, B. (2005, June). *The role of floor control and of ontology in argumentative activities with discussion-based tools*. Paper presented at the Computer-supported collaborative learning conference, Taipei, Taiwan.

Glassner, A., Weinstock, M., & Neuman, Y. (2005). Pupils' evaluation and generation of evidence and explanation in argumentation. *British Journal of Educational Psychology, 75*, 105–118.

Graff, G. (2003). *Clueless in academe: How schooling obscures the life of the mind*. New Haven, CT: Yale University Press.

Hersh, R. (2005, November). What does college teach? *Atlantic Monthly*, 140–143.

Hofer, B., & Pintrich, P. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research, 67*, 88–140.

Hofer, B. & Pintrich, P. (Eds.). (2002). *Epistemology: The psychology of beliefs about knowledge and knowing*. Mahwah, NJ: Erlbaum.

Kuhn, D. (1991). *The skills of argument*. Cambridge, UK: Cambridge University Press.

Kuhn, D. (1995). Microgenetic study of change: What has it told us? *Psychological Science, 6*, 133–139.

Kuhn, D. (2005). *Education for thinking*. Cambridge, MA: Harvard University Press.

Kuhn, D., & Franklin, S. (2006). The second decade: What develops (and how)? In W. Damon & R. Lerner (Series Eds.) & D. Kuhn & R. Siegler (Vol. Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language.* (6th ed., pp. 953–993). Hoboken, NJ: Wiley.

Kuhn, D., & Park, S. (2005). Epistemological understanding and the development of intellectual values. *International Journal of Educational Research, 43*, 111–124.

Kuhn, D., Shaw, V., & Felton, M. (1997). Effects of dyadic interaction on argumentive reasoning. *Cognition and Instruction, 15*, 287–315.

Kuhn, D., & Udell, W. (2003). The development of argument skills. *Child Development, 74*, 1245–1260.

Kuhn, D., & Udell, W. (2007). Coordinating own and other perspectives in argument. *Thinking and Reasoning, 13*, 90–104.

Leitao, S. (2003). Evaluating and selecting counterarguments. *Written Communication, 20*, 269–306.

Means, M., & Voss, J. (1996). Who reasons well? Two studies of informal reasoning among students of different grade, ability, and knowledge levels. *Cognition and Instruction, 14*, 139–178.

Miller, P. H., & Coyle, T. R. (1999). Developmental change: Lessons from microgenesis. In E. K. Scholnick, K. Nelson, S. A. Gelman, & P. H. Miller (Eds.), *Conceptual development: Piaget's legacy* (pp. 209–239). Mahwah, NJ: Erlbaum.

Moshman, D. (2005). *Adolescent psychological development: Rationality, morality, and identity* (2nd ed.). Mahwah, NJ: Erlbaum.

Nussbaum, E. M. (2003). Appropriate appropriation: Functionality of student arguments and support requests during small-group classroom discussions. *Journal of Literacy Research, 34*, 501–544.

Nussbaum, E. M. (2005). The effects of goal instructions and need for cognition on interactive argumentation. *Contemporary Educational Psychology, 30*, 286–313.

Nussbaum, E. M., & Kardash, C. (2005). The effects of goal instructions and text on the generation of counterarguments during writing. *Journal of Educational Psychology, 97*, 157–169.

Nussbaum, E. M., & Sinatra, G. (2003). Argument and conceptual engagement. *Contemporary Educational Psychology, 28*, 384–395.

Perkins, D. (1985). Postprimary education has little impact on informal reasoning. *Journal of Educational Psychology, 77*, 562–571.

Reznitskaya, A., Anderson, R., McNurlen, B., Nguyen-Jahiel, K., Archodidou, A., & Kim, S. (2001). Influence of oral discussion on written argument. *Discourse Processes, 32*, 155–175.

Schwarz, B., Newman, Y., & Biezuner, S. (2000). Two wrongs may make a right ... If they argue together. *Cognition and Instruction, 18*, 461–494.

Siegler, R. (2006). Microgenetic studies of learning. In W. Damon & R. Lerner (Series Eds.) & D. Kuhn & R. Siegler (Vol. Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language* (6th ed., pp. 464–510). Wiley.

Smetana, J., Killen, M., & Turiel, E. (1991). Children's reasoning about interpersonal and moral conflicts. *Child Development, 62*, 629–644.

Stanovich, K. (1999). *Who is rational? Studies of individual differences in reasoning*. Mahwah, NJ: Erlbaum.

Udell, W. (2007). Enhancing adolescent girls' argument skills in reasoning about personal and non-personal decisions. *Cognitive Development, 22*, 341–352.

Voss, J. (Ed.). (2001). Argumentation in psychology. *Discourse Processes, 32*(2 & 3).

Voss, J., & Means, M. (1991). Learning to reason via instruction in argumentation. *Learning and Instruction, 1*, 337–350.

Wainryb, C., Shaw, L., Laupa, M., & Smith, K. (2001). Children's, adolescents', and young adults' thinking about different types of disagreements. *Developmental Psychology, 37*, 373–386.

Walton, D. N. (1989). Dialogue theory for critical thinking. *Argumentation, 3*, 169–184.

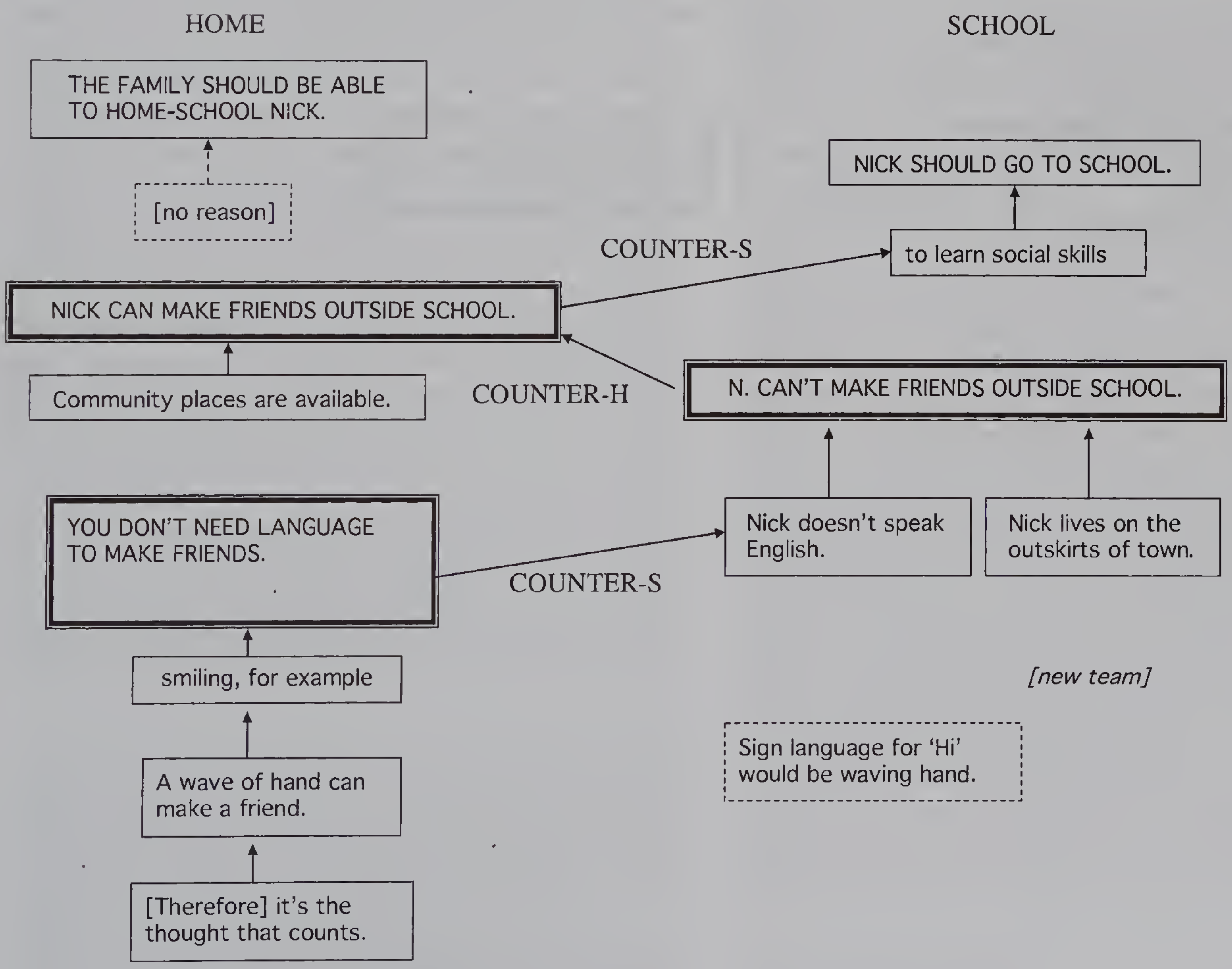
Weinstock, M., Neuman, Y., & Tabak, I. (2004). Missing the point or missing the norms? Epistemological norms as predictors of students' ability to identify fallacious arguments. *Contemporary Educational Psychology, 29*, 77–94.

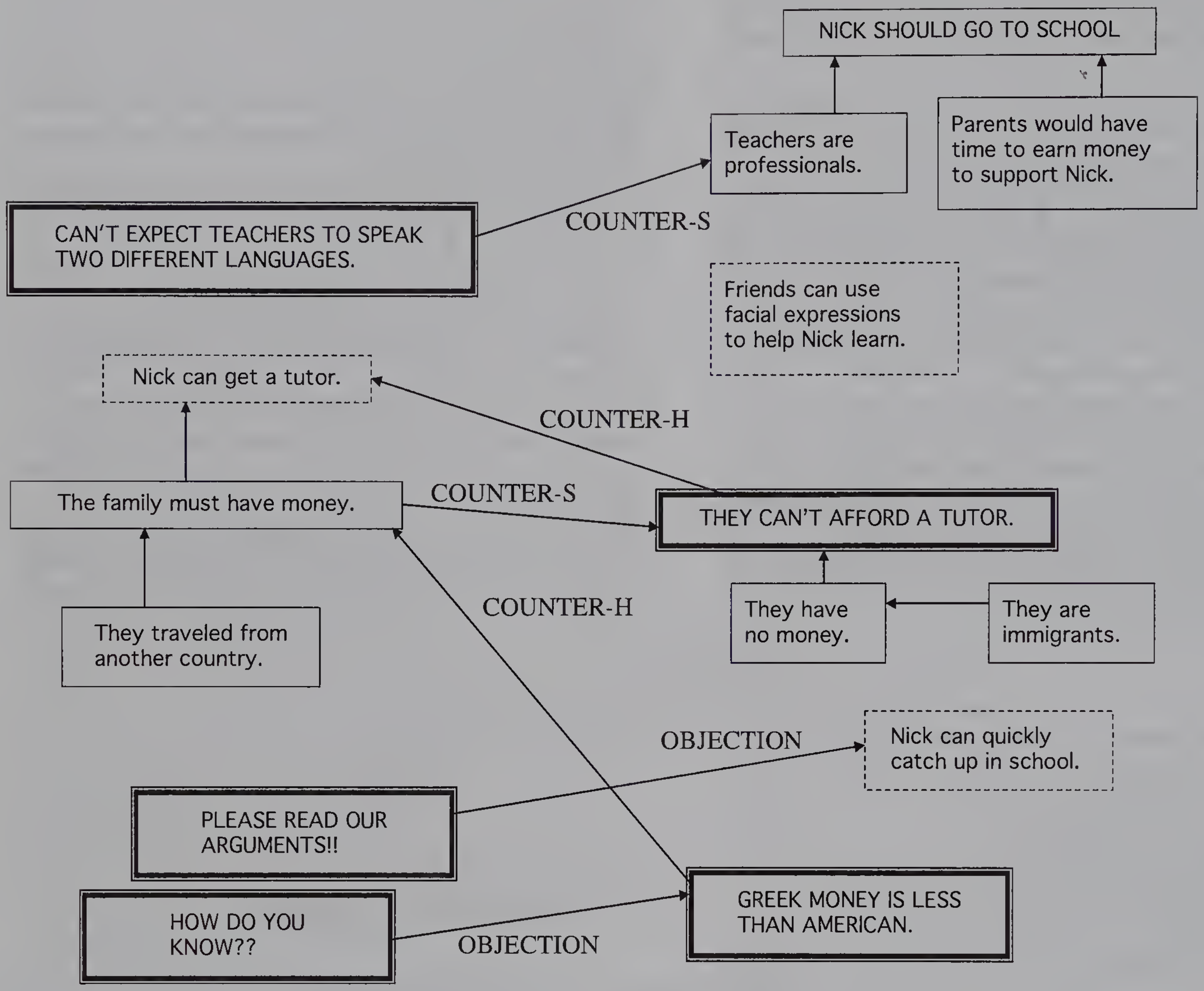
Wiley, J., & Voss, J. (1999). Constructing arguments from multiple sources: Tasks that promote understanding and not just memory for text. *Journal of Educational Psychology, 91*, 301–311.

Zohar, A., & Nemet, F. (2002). Fostering students' knowledge and argumentation skills through dilemmas in human genetics. *Journal of Research in Science Teaching, 39*, 35–62.

Appendix

Group Argument Map Produced at Homeschool Showdown


HOME
SCHOOL
THE FAMILY SHOULD BE ABLE TO HOME-SCHOOL NICK.
[no reason]
NICK SHOULD GO TO SCHOOL.
COUNTER-S
to learn social skills
NICK CAN MAKE FRIENDS OUTSIDE SCHOOL.
Community places are available.
COUNTER-H
N. CAN'T MAKE FRIENDS OUTSIDE SCHOOL.
YOU DON'T NEED LANGUAGE TO MAKE FRIENDS.
Nick doesn't speak English.
Nick lives on the outskirts of town.
COUNTER-S
smiling, for example
[new team]
A wave of hand can make a friend.
Sign language for 'Hi' would be waving hand.
[Therefore] it's the thought that counts.


NICK SHOULD GO TO SCHOOL
Teachers are professionals.
Parents would have time to earn money to support Nick.
CAN'T EXPECT TEACHERS TO SPEAK TWO DIFFERENT LANGUAGES.
COUNTER-S
Friends can use facial expressions to help Nick learn.
Nick can get a tutor.
COUNTER-H
The family must have money.
COUNTER-S
THEY CAN'T AFFORD A TUTOR.
They traveled from another country.
COUNTER-H
They have no money.
They are immigrants.
OBJECTION
Nick can quickly catch up in school.
PLEASE READ OUR ARGUMENTS!!
HOW DO YOU KNOW??
OBJECTION
GREEK MONEY IS LESS THAN AMERICAN.

Child Development, September/October 2008, Volume 79, Number 5, Pages 1329 – 1343

The Influence of Race and Gender on Children's Conversations and Playmate Choices

Patrick J. Leman
Royal Holloway, University of London

Virginia L. Lam
University of East London

The present study examined the influence of race and gender on children's conversations and friendship choices. Four hundred and twenty-eight children (*M* age = 7.5 years, *SD* = 0.34) from 2 racial minority groups (i.e., African Caribbean and South Asian) and the racial majority group (i.e., European) chose a picture of a playmate together with a peer. Race influenced the levels of assertion and affiliation in children's conversations. The effects of race on conversation also varied according to the gender of the children involved in interaction. Same-race pairs tended to choose in-group playmates, but same-race minority pairs showed less marked in-group preference. Cross-race pairs selected a majority-group child as a playmate most often.

The present study investigated, for the first time, how race influences the social dynamics of conversations between children when they choose a playmate together. Previous research has established an association between race and friendship from preschool onward. A consistent finding is that there is racial cleavage in playmate choices; that is, children tend to play with peers from the same-race group more frequently than with peers from other-race groups (Finkelstein & Haskins, 1983; Fishbein & Imai, 1993). Although these in-group racial biases can be identified in the attitudes, judgments, and behavior that children exhibit, a key context in which these biases become visible is in interaction and conversation. Although conversations with adults may not be free from the influence of race, peer conversation is where most children will establish their friendships and learn about relationships with others for themselves. By focusing on the influence of race on conversations and playmate choices between peers, the present study provides a fresh perspective on questions of children's racial and intergroup understanding and on how aspects of children's identity may influence their conversations. Moreover, this research contributes to a better understanding of the links between race and other related processes including friendship (Aboud, Mendelson, & Purdy, 2003; Lease & Blake, 2005), peer relationships (Killen & Stangor, 2001; Wright & Tropp, 2005), and the reduction of prejudice through intergroup contact (Aboud & Fenwick, 1999; Doyle & Aboud, 1995; Kenworthy, Turner, Hewstone, & Voci, 2005). Three questions guided the present study: First, does race influence the social dynamics of children's conversations about who to choose as a playmate? Second, does race influence the joint choices of playmates by children? Third, what are the relations between children's conversations and their choices of playmates?

The focus on the specifics of children's exchanges offers an opportunity to examine how contact can be productive or destructive in terms of broader relationships between children. Interpersonal contact is often proposed as a mechanism for reducing prejudice and creating good race relations (Cameron, Rutland, Brown, & Douch, 2006). However, if this contact creates or maintains negative forms of interaction, or reinforces stereotypes, its value is questionable. Alongside informing debates about interracial contact between children, the present study also addresses a gap in research on children's conversations because previous studies in this area have focused either on racially homogenous samples or have not directly examined the impact of race on conversations (Leaper & Smith, 2004; Quintana et al., 2006).

It is, perhaps, surprising that no previous work has sought to unravel the ways in which a child's race can affect their conversation style. In contrast, a good deal of research has charted how gender influences

---

The research reported here was supported by a grant from the Nuffield Foundation (SGS/00735/G). We would like to thank all the schools, teachers, and children who participated in this research. We would also like to thank Martyn Barrett, Margaret Harris, Carol Levine, Katy Marcus, Ned Palmer, Kathy Rastle, Harriet Tenenbaum, Dawn Watling, Donald Wells, and five reviewers for feedback on earlier drafts of this article.

Correspondence concerning this article should be addressed to Patrick J. Leman, Department of Psychology, Royal Holloway, University of London, Egham, Surrey TW20 0EX, United Kingdom. Electronic mail may be sent to Patrick.Leman@rhul.ac.uk.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7905-0008

children's conversations. Gender differences in conversation have been attributed to different socialization experiences that result in separate male and female (or two cultures of) conversation styles (Maltz & Borker, 1982). As with gender, from early childhood, children realize there are different races (Aboud, 1988), and by the end of middle childhood, children associate different racial groups with specific stereotypes and make predictions about behavior and preferences that are based on race (Lam & Leman, 2003; Powlishta, Serbin, Doyle, & White, 1994). In this respect, children develop a sense of their own and others' identities as a result of their membership in different social groups (Tajfel & Turner, 1986). Membership in these groups entails different socialization experiences that may feed into race differences in conversational style and dynamics. Much like the two cultures approach to gender and conversation style, there may be different cultures of conversation style associated with race.

Gender differences in conversation have also been attributed to status differences or status characteristics that stem from perceived social asymmetries in gender relations (James & Drakich, 1993). In this respect, cross-gender interaction (i.e., between a boy and a girl) involves an asymmetric status relation between children because the contact is with a member of the gender out-group. However, interaction exclusively between children of the same gender involves contact with another in-group member and, as a consequence, may be more symmetric. These in-group and out-group dynamics are associated with differences in status that, it is argued, lead to differences in conversational dynamics between boys and girls in same- and cross-gender interaction. Race, like gender, marks out social groups that are associated with different status characteristics. Given that gender is a dominant force in shaping children's conversation styles, other aspects of a child's social identity such as race may influence conversation styles too.

The first set of hypotheses addressed whether race influences children's conversation dynamics. Our first hypothesis directly examined whether there would be differences in conversational style between race groups. Specifically, we predicted that there would be differences in children's levels of assertion and affiliation depending on a speaker's racial group (i.e., African Caribbean, European, and South Asian). Assertion and affiliation are identifiable and recognizable features of children's conversations (Leaper, 2000) that offer a systematic means for gauging the dimensions of conflict, cooperation, and dominance in children's interactions. As such, they are markers of the social dynamics of children's conversations and have been linked to gendered styles of speech and to differences in status between children. Although we predicted that race would affect children's conversation styles, in the absence of previous research, we made no predictions about how the levels of assertion and affiliation relate to particular race groups.

We also anticipated that children would respond to the race of their conversation partner and the racial mix of a pair because, like gender, race is a social category associated with differences in status between groups. European (i.e., majority group) children and adults are often perceived to hold higher status in society (Brown & Bigler, 2005). Therefore, our second hypothesis examined whether status characteristics would affect the level of assertion and affiliation displayed by children. We predicted that in cross-race pairs, European children would be more assertive than their African Caribbean and South Asian conversation partners. Our third hypothesis was derived from work on intergroup dynamics indicating that individuals tend to affiliate more with in-group members and disaffiliate (or, at least, do not affiliate as strongly) with out-group members. We therefore predicted that children in same-race pairs would display higher levels of affiliation than children in cross-race pairs.

Our primary focus was on race and conversation, but a corollary issue is whether gender and race interact to affect children's conversations in different ways. No previous work has explored directly the connection among race, gender, and conversation. Some work has suggested that some elements of the influence of gender on conversation may cut across racial groups. Leaper, Tenenbaum, and Shaffer (1999) examined the conversations of African American children using a puppet-play task. The results from that study broadly replicated those from an earlier study (Leaper, 1991) of mostly European American children that examined whether or not gender influenced children's conversations. Both studies found that conversations between boys were more domineering, whereas conversations between girls were generally more collaborative. Results also showed that boys and girls were less controlling and domineering when in conversation with a girl than when in conversation with a boy. Indeed, a meta-analysis of studies of gender and conversation revealed that, on average, girls are more likely to use language to connect or affiliate with others, whereas boys are more likely to use language to dominate or assert themselves over a conversation partner (Leaper & Smith, 2004). Subsequent to our analysis of race, in the present study, we analyzed boys' and girls' pairs separately to explore whether these effects of gender in conversation interact with those of race.

The second set of hypotheses addressed whether race influences children's choice of a playmate. Racial cleavage in play is evident from preschool (Fishbein & Imai, 1993; Hallinan & Teixeira, 1987a; Howes & Wu, 1990), but the divisions between children from different racial groups are even more apparent among older children (Boulton & Smith, 1992). More recent research has indicated that the number of cross-race friendships declines with age and that such friendships are less stable and intimate than same-race friendships. European American children are less likely than minority-group children to choose a peer from another racial group as a friend (Aboud et al., 2003), but the racial mix of a school affects the level of minority-group bias and friendship choices among European American children (McGlothlin & Killen, 2006; McGlothlin, Killen, & Edmonds, 2005). Additionally, majority and minority children who have friends from ethnic minority groups in a school differ from one another on a number of social, behavioral, and affective measures (Lease & Blake, 2005). Therefore, although evidence suggests that race affects children's friendship choices, the association between race and friendship is complex and can be influenced by context. By exploring playmate choices in a conversational context, the present study provides a fresh perspective on how race influences children's social relationships.

Our fourth hypothesis reflects the broad finding that children tend to choose playmates from their own race. We predicted that same-race pairs would choose playmates reflecting an in-group preference. The fifth hypothesis predicted that in pairs from different race groups (i.e., cross-race pairs), there would be less agreement about the choice of a preferred playmate. This prediction is consistent with previous research with adults (Garcia-Marques & Mackie, 2001), which has found that individuals are more easily persuaded by an in-group than an out-group member.

The third set of hypotheses concerns the links among race, conversation, and children's choices of a playmate. Contact between individuals of different groups can reduce prejudice toward out-groups in general (e.g., Pettigrew, 1997; Pettigrew & Tropp, 2006) and in minority groups in particular (Tropp & Pettigrew, 2005), although the type of contact is crucial in determining its effectiveness in reducing prejudice (Kenworthy et al., 2005). Conversation is a key element in any contact. Aboud and Levy (2000) found that conversations between peers reduced racial prejudice and promoted positive racial attitudes in children. Aboud and Doyle (1996) found that children who talked with a less prejudiced peer experienced a reduced level of prejudice (see also Aboud & Fenwick, 1999).

If race affects children's conversation style, then it is also likely that race will affect the outcomes or decisions that are made. Leman, Ahmed, and Ozarow (2005) found that levels of assertion and affiliation correlated with different outcomes in children's conversations. For example, compared to conversations where one child won, when children reached a compromise, conversations tended to be more balanced, both in terms of each partner's contribution to the discussion and their levels of assertion and affiliation. Consequently, our sixth hypothesis predicted that unresolved conversations (i.e., where there is no agreement about a playmate) would involve high levels of assertion and low levels of affiliation.

The seventh and eighth hypotheses stem from an assumption that race and conversation dynamics would link with the decisions children reached about a playmate. Our seventh hypothesis predicted that when children in same-race pairs agree on a playmate from their racial in-group, they will exhibit high levels of affiliation. In contrast, for our eighth hypothesis, we predicted that when children in these pairs choose a playmate from their racial out-group, they will exhibit low levels of affiliation. Although we anticipated fewer resolved conversations in cross-race pairs, we made no predictions about the links between conversation measures and choices of playmate when cross-race pairs agreed.

The focus on children between 7 and 8 years of age identifies a particularly important age group for the study of racial understanding. Although children hold negative racial attitudes toward out-groups during early childhood, these attitudes start to become more positive from around 7 years, at least in terms of their explicit expression. This increasing positivity in attitudes, however, coincides with a decline in the selection of children from other racial groups as friends (Hallinan & Teixeira, 1987a, 1987b). Perhaps from this age peer influence becomes a more powerful force in children's decision making and contributes to in-group (i.e., same-race) friendship choices. Therefore, although at this age children's attitudes may be becoming more positive toward minority groups, increasing peer (i.e., in-group) influence may limit the number of cross-race friends (Abrams, Rutland, Cameron, & Marques, 2003). These changes may, in turn, be linked to more general developmental changes in underlying cognitive processes (see Powlishta et al., 1994). Participating children were members of the predominant racial groups in London, which constitute the majority group and the two largest racial minority groups in both the city and the country as a whole. This choice of racial groups made sense not only because it allowed a maximum number of interactions to be observed but also because such interactions between children would be most frequent in this multiethnic context.

Method

Participants

Four hundred and twenty-eight children (*M* age = 7.5 years, *SD* = 0.34, 210 boys and 218 girls) participated in the present study. Children were recruited from eight schools from the same working class (i.e., low-socioeconomic-status) area of London, United Kingdom. All the schools were in close proximity to one another and shared a similar racial composition. There were 103 children of South Asian racial origin who include those with familial racial origins in India, Pakistan, and Bangladesh. There were 99 children of African and Caribbean racial origin. Finally, there were 226 children of European racial origin. Mixed-race children and those from other ethnic groups participated in paired discussions, but because the numbers in each group were small, these data are not included in the analysis.

Schools keep no record of pupils' race, but the most recent data (United Kingdom Government, Office for National Statistics, 2005) for the racial composition of residents in this part of London indicated that 64% of the population was European, 11.2% was South Asian, and 20.4% was African Caribbean. Nationally, 92.1% of the population is European (i.e., the largest racial group), 4% is South Asian (i.e., the second largest group), and 2% is African Caribbean (i.e., the third largest group). However, teachers estimated that between 65% and 75% of the students came from ethnic minority groups, which is plausible given the high birth rates among ethnic minority groups. In other words, although African Caribbean and South Asian children constitute a racial minority across the United Kingdom, there was no clear racial majority group in the schools that participated in this study.

To facilitate analysis, pairs were divided into two separate samples: the first sample (*N* = 220) compared South Asian and European children and the second sample (*N* = 208) compared African Caribbean and European children. The two samples produced two separate sets of all-European pairs. A comparison of these two sets enabled a cross-check for equivalence of these all-European pairs from the two samples.

Procedure

Eight photographs, each showing a target child who was in the same age group as the participants, were presented to the pairs. For the European and South Asian sample, four of the eight target children were European (i.e., two boys and two girls) and four of the eight target children were South Asian (i.e., two boys and two girls). For the European and African Caribbean sample, four of the eight target children were European (using the same photos shown to the European and South Asian sample) and four of the eight target children were African Caribbean (i.e., two boys and two girls). The eight target children's ages, attractiveness, and racial typicality were matched by adult and child judges in testing before the interaction phase of the study.

Children were placed in a pair with a peer of the same sex. There were five different pairings for boys and girls: (a) all African Caribbean (henceforth AC–AC), (b) all South Asian (henceforth SA–SA), (c) all European (henceforth Eu–Eu), (d) African and European (henceforth AC–Eu), and (e) Asian and European (henceforth SA–Eu). Because of insufficient numbers of racial minority-group children, it was not possible to create African and Asian pairings. Only same-sex pairs were used in this study because including boy–girl pairs would have added greatly to the complexity of the design. Children were paired to ensure an equal number of different pair types. Dyad composition was also determined after discussing the children's verbal ability, intelligence, and sociometric status with the class teacher. The latter matching variable meant that the dyadic partners were seen to be reasonably agreeable classmates who were not best friends. Pairs were not explicitly matched for socioeconomic status, but given the area in which the schools were located, the overwhelming majority of children were from low-socioeconomic backgrounds.

Each pair of children was interviewed in a quiet room by a female interviewer who said, "I am interested in how children choose other children to play with." The photographs of the potential playmates were laid out in random order in front of the pair, and each pair was asked, "Imagine that these children were going to be new children in your class. Now, if you had to pick just one of them you both would most want to play with, which would that be? Do think about it together before picking." Each dyad was permitted 3 min to reach their decision. Pairs were able to choose one photograph from the eight. In the event that the children did not reach a consensus at the end of 2 min, the children were told they had 1 more minute to decide. The interviewer's input was kept to a minimum unless the children initiated interaction (e.g., asking questions) or were silent. If the children were silent, the interviewer attempted to encourage the children to interact. A video camera was used to film the interaction between the dyadic partners.

There were five possible outcomes in the European and South Asian sample: European boy, European girl, South Asian boy, South Asian girl, and no decision (unresolved). There were also five possible outcomes for the European and African sample: European boy, European girl, African boy, African girl, and no decision (unresolved).

Given there were multiple racial pairings, it would have been impractical to counterbalance interviewer race and gender across all pairings. The interviewer's racial origin, however, was East Asian (i.e., Chinese), which constituted an out-group for all the children in this study. Therefore, many of the potential pitfalls of having an interviewer from the majority group were avoided. In addition, using a single interviewer ensured consistency of practice across pairings.

Measures of Conversational Style

The Psychosocial Processes Ratings Scale (PPRS; for more details, see Leaper, 2000) was used in the present study to explore children's use of assertion and affiliation. PPRS has two separate, independent 7-point scales for rating levels of personal assertion and interpersonal affiliation. These scales provide detailed descriptions about specific, common behaviors, such as asking for another's opinion (Level 6 affiliation, moderately interdependent) or agreeing with another's position without comment (Level 2 assertion, moderately nondirect). The first author (of European ethnic origin) of this article made separate ratings for each child every 5 s. Therefore, there are ratings every 5 s for a child's behavior, and these ratings range from unassertive (i.e., nondirective behavior such as sitting passively and withdrawing from conversation) to highly assertive behavior (i.e., giving commands or demands, strong disagreement, and aggression). Similarly, there were ratings every 5 s for levels of affiliation, and these ratings ranged from unaffiliative (i.e., highly distant behavior, such as active resistance, outright rejection, or strong disagreement) to highly interdependent behavior (i.e., joining in or clear cooperative activity). Average assertion and affiliation scores were calculated for each child. High scores on each of the separate scales indicated high assertion and high affiliation.

Reliability of Conversation Measures

To test the reliability of the PPRS scores, a second judge (i.e., the second author, who is of Chinese ethnic origin) blind coded 20% (i.e., 42) of the conversations that were selected at random. Agreement on coding for PPRS categories was good: assertion, $\kappa = .84$; affiliation, $\kappa = .72$.

Results

Design and Analysis

A between-subjects design was used to analyze each sample. The child's race (i.e., African Caribbean, European, and South Asian) was a blocking variable, and the predictor variable was the racial mix of the pairing. For the South Asian and European sample, the racial mix was either two South Asian children (SA–SA, $N = 35$), two European children (Eu–Eu, $N = 42$), or one South Asian child and one European child (SA–Eu, $N = 33$). For the African Caribbean and European sample, the racial mix was either two African Caribbean children (AC–AC, $N = 32$), two European children (Eu–Eu, $N = 37$), or one African Caribbean child and one European child (AC–Eu, $N = 35$).

There were three dependent variables and measures. Two dependent variables related to conversation measures that concerned how the racial mix of the pair and the race of the speaker and partner influenced children's use of communication styles (i.e., assertion and affiliation). Analyses of children's levels of assertion and affiliation were based on scores that ranged from 1 to 7. The joint decisions that children in different pairs reached about which playmate they would pick related to the third dependent variable. All the children chose a same-gender playmate; therefore, this outcome choice of preferred playmate was classified only in terms of race (i.e., African Caribbean, European, and South Asian).

The results reported here focus on three separate units of analysis: (a) the pair, (b) the individual child, and (c) the child's conversation partner. The analysis of dyadic interaction presents a particular statistical challenge to researchers. Specifically, the contribution to a conversation of one dyad member is (almost inevitably) dependent on the contribution of their partner (Gonzalez & Griffin, 2000). Correcting for nonindependence allows an assessment of the impact of a speaker's and a partner's race on a conversation. It also allows a comparison of the extent of race differences between same-race and cross-race pairs.

In the present study, we adapted a technique developed by Carli (1989a, 1989b) and Kraemer and Jacklin (1979) for use in analyzing gender effects in conversation but replaced the binary distinction between gender groups with a binary distinction between different racial groups in each sample. This technique involves applying a set of formulas to the means and mean error terms in two separate analyses of variance (ANOVAs). In this study, a first ANOVA with race as a between-groups factor was carried out for the same-race pairs, and a second ANOVA with race as a within-groups factor was carried out for the cross-race pairs.

The formulas for each comparison share a common denominator: $([2M_{se} + 2MS_e'][1/n])^{1/2}$, where M_{se} equals mean error score for race effects in same-race pairs (in each separate group), MS_e' equals mean error score for race effects in cross-race pairs (in each separate sample), and n equals the number of observations used in the analysis. Using the South Asian–European group as an example, the numerators for each effect (i.e., speaker, partner, and interaction) were as follows: speaker race effects ($[M_{\text{as}} + M_{\text{am}}] - [M_{\text{ws}} - M_{\text{wm}}]$), partner race effects ($[M_{\text{as}} - M_{\text{am}}] + [M_{\text{wm}} - M_{\text{ws}}]$), and interaction effect ($[M_{\text{as}} - M_{\text{am}}] - [M_{\text{wm}} + M_{\text{ws}}]$), where M_{as} equals mean for Asian in same-race pairs (SA–SA), M_{am} equals mean for South Asian in cross-race pairs (SA–Eu), M_{ws} equals mean for European in same-race pairs (Eu–Eu), and M_{wm} equals mean for European in cross-race pairs (SA–Eu).

These formulas transform the ANOVA output scores and provide separate values of the t statistic for tests that relate to race in each of the separate parameters (i.e., speaker and partner race effects and interaction effects). For speaker and partner race effects, a positive value of t indicated a larger score for South Asian or African Caribbean children compared to European children. For the interaction effect, a positive value of t indicated that the difference in the use of conversation measures is greater in same-race pairs than in cross-race pairs.

Race and Conversation Style

The first set of hypotheses concerned the links between race and conversation. Our first hypothesis predicted differences in levels of assertion and affiliation that were linked to a speaker's race. Our second hypothesis predicted that European children would be more assertive than their African Caribbean and South Asian conversation partners in cross-race pairs. Our third hypothesis predicted that children in same-race pairs would display higher levels of affiliation than children in cross-race pairs. The results relating to these three hypotheses are organized in terms of effects of speaker race, effects of partner race, and the interaction effect. The analyses are presented first for all children in a sample, then separately for boys and girls in the sample.

Variations in the use of assertion and affiliation were examined for both samples. Speaker, partner, and interaction effects for the South Asian and European sample are presented in Table 1. Corresponding mean scores and standard deviations are broken down by race and pair type (same- and cross-race pairs) in Table 2. Additionally, results from the between-group (comparing between South Asian and European children in same-race pairs) and within-group (comparing South Asian and European children in the cross-race pairs) ANOVAs are given in Table 2. Table 3 reports speaker, partner, and interaction effects for the African Caribbean and European sample. Table 4 displays mean scores, standard deviations, and associated results from the between- and within-group ANOVAs for this sample.

Speaker race, partner race, and interaction effects. As predicted, there were significant differences in the use of assertion and affiliation associated with speaker race. European children were more assertive than both South Asian and African Caribbean children. Overall, South Asian children were more affiliative than European children, although there were no differences in levels of affiliation between African Caribbean and European children.

Tables 1 and 3 also indicate that there were effects of partner race on use of assertion and affiliation. In both samples, children use more affiliation when speaking with a partner who is South Asian and

Table 1

Values of t *and Effect Size (Cohen's* d*) for Speaker Race, Partner Race, and Interaction Effects Associated With the Use of Conversational Measures for South Asian and European Children*

	Speaker race		Partner race		Interaction	
	t	*d*	*t*	*d*	*t*	*d*
All children						
Assertion	−1.98*	−0.27	4.06***	0.55	0.19	0.03
Affiliation	4.44***	0.60	1.96*	0.24	6.00***	0.81
Boys only						
Assertion	−6.71***	−1.29	2.10*	0.40	−3.20***	−0.62
Affiliation	0.05	0.01	1.59	0.31	4.59***	0.88
Girls only						
Assertion	1.49	0.29	−3.09**	−0.59	−2.55**	−0.49
Affiliation	2.30*	0.44	0.31	0.06	3.50***	0.67

Note. For speaker and partner race effects, a positive value of t indicates a higher score for Asian than European children. A positive value of t in the interaction effect reflects a greater difference in same-race compared with cross-race pairs. The degrees of freedom for comparisons of speaker race, partner race, and interaction for the t tests are as follows: all children ($df = 218$), boys only ($df = 108$), girls only ($df = 108$).
$^*p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Table 2
Mean Use of Conversational Measures and Associated Univariate Statistics for the South Asian and European Sample

	Same-race pairs							Cross-race pairs						
	South Asian (SA – SA)		European (Eu – Eu)					South Asian (in SA – Eu)		European (in SA – Eu)				
	M	*SD*	*M*	*SD*	*F*	*df*	η^2	*M*	*SD*	*M*	*SD*	*F*	*df*	η^2
All children														
Assertion	5.13	1.34	5.45	1.31	2.13	1, 152	.01	5.23	1.18	5.34	1.31	0.13	1, 64	.00
Affiliation	4.22	1.37	3.91	1.25	2.19	1, 152	.01	3.60	1.20	3.61	1.35	0.00	1, 64	.00
Boys														
Assertion	5.00	1.32	5.42	1.17	2.27	1, 78	.03	5.08	0.97	5.95	0.57	9.05*	1, 28	.24
Affiliation	4.43	1.40	4.13	1.24	1.06	1, 78	.01	3.64	1.19	3.94	1.19	0.46	1, 28	.02
Girls														
Assertion	5.29	1.37	5.47	1.45	0.30	1, 72	.00	5.35	1.35	4.83	1.53	1.18	1, 34	.03
Affiliation	3.92	1.32	3.69	1.25	0.90	1, 72	.01	3.56	1.24	3.34	1.15	0.23	1, 34	.01

Note. The means and associated statistics compare first race effects between same-race pairs (two South Asian children in a pair, SA – SA, and two European children in a pair, Eu – Eu) and then race effects between children in cross-race pairs (i.e., the South Asian child and the European child in SA – Eu). The degrees of freedom for the *t* tests are computed from the number of observations on which the means are based. In this study, that includes children in all-South Asian pairs (SA – SA, $n = 70$), children in all-European pairs (Eu – Eu, $n = 84$), South Asian children in SA – Eu ($n = 33$), and European in SA – Eu ($n = 33$). Proportion scores do not sum to 1.00 because some children offered no justifications and scored 0.
$^*p < .05$.

African Caribbean than when speaking with a partner who is European. Children also used more assertion when speaking with a South Asian partner compared with a European partner.

European children were more assertive than their African Caribbean partners in cross-race pairs, as indicated by a negative value for the associated interaction effect (see Table 3). However, there was no similar interaction effect among the South Asian and European sample. Levels of affiliation were higher in same-race than in cross-race pairs (see Tables 2 and 4). Additionally, race differences in the use of affiliation were greater comparing between same-race pairs than between each child in a cross-race pair.

Table 3
Values of t *and Effect Size (Cohen's* d*) for Speaker Race, Partner Race, and Interaction Effects Associated With the Use of Conversational Measures for African Caribbean and European Children*

	Speaker race		Partner race		Interaction	
	t	*d*	*t*	*d*	*t*	*d*
All children						
Assertion	−2.92**	−0.41	−0.28	−0.04	−2.87**	−0.40
Affiliation	1.55	0.22	4.13***	0.58	4.58***	0.64
Boys only						
Assertion	−0.09	−0.02	2.80**	0.57	−3.63***	−0.73
Affiliation	−0.72	−0.15	−0.48	−0.10	2.46**	0.50
Girls only						
Assertion	−1.49	−0.29	−0.42	−0.08	−0.76	−0.15
Affiliation	3.29***	0.64	6.48***	1.26	3.28***	0.64

Note. For speaker and partner race effects, a positive value of *t* indicates a higher score for African Caribbean than European children. A positive value of *t* in the interaction effect reflects a greater difference in same-race compared with cross-race pairs. The degrees of freedom for comparisons of speaker race, partner race, and interaction for the *t* tests are as follows: all children ($df = 204$), boys only ($df = 98$), girls only ($df = 106$).
$^{**}p < .01$. $^{***}p < .001$.

Assertion in cross-race pairs. To test our second hypothesis, we conducted independent *t* tests to compare the levels of assertion for each child in the cross-race pairs in each sample (i.e., SA – Eu and AC – Eu). Although mean scores in Tables 2 and 4 indicate that in cross-race pairs European children used more assertion than either their South Asian or African Caribbean partners, these differences were not significant.

Affiliation in same- versus cross-race pairs. The third hypothesis was that levels of affiliation would be higher in same- than in cross-race pairs. A one-way ANOVA was conducted for each sample comparing mean levels of affiliation in each pair type. For both samples, there was a small but significant effect of pair type on use of affiliation. In the South Asian and European sample, $F(2, 217) = 3.86$, $p < .05$, $\eta^2 = .03$, post hoc Tukey's honestly significant difference (HSD) tests ($p < .05$) indicated that there was significantly less affiliation in the cross-race pair (SA – Eu) compared with the all-South Asian pair (SA – SA). In the African Caribbean and European sample, $F(2, 205) =$

Table 4
Mean Use of Conversational Measures and Associated Univariate Statistics for the African Caribbean and European Sample

	Same-race pairs							Cross-race pairs						
	African Caribbean (AC–AC)		European (Eu–Eu)					African Caribbean (in AC–Eu)		European (in AC–Eu)				
	M	*SD*	*M*	*SD*	*F*	*df*	η^2	*M*	*SD*	*M*	*SD*	*F*	*df*	η^2
All children														
Assertion	4.85	1.25	5.10	1.18	1.42	1, 136	.01	5.10	1.22	5.30	1.28	.47	1, 68	.01
Affiliation	4.28	1.46	3.83	1.25	3.84†	1, 136	.03	3.59	1.60	3.80	1.54	.30	1, 68	.00
Boys														
Assertion	4.84	1.02	5.13	1.02	1.30	1, 66	.02	5.22	1.19	5.53	1.37	.46	1, 30	.02
Affiliation	4.18	1.52	4.33	1.11	0.24	1, 66	.00	3.93	1.88	3.96	1.84	.00	1, 30	.00
Girls														
Assertion	4.87	1.44	5.08	1.35	0.40	1, 68	.01	5.00	1.27	5.11	1.20	.09	1, 36	.00
Affiliation	4.38	1.41	3.31	1.18	11.91*	1, 68	.15	3.31	1.30	3.66	1.28	.70	1, 36	.02

Note. The means and associated statistics compare first race effects between same-race pairs (two African children in a pair, AC – AC, and two European children in a pair, Eu – Eu) and then race effects between children in cross-race pairs (i.e., the African child and the European child in AC – Eu). The degrees of freedom for the *t* tests are computed from the number of observations on which the means are based. In this study, that includes children in all-African pairs (AC – AC, $n = 64$), children in all-European pairs (Eu – Eu, $n = 74$), African children in AC – Eu ($n = 35$), and European in AC – Eu ($n = 35$). Proportion scores do not sum to 1.00 because some children offered no justifications and scored 0. $^\dagger p < .10$. $^* p < .05$.

3.11, $p < .05$, $\eta^2 = .03$, post hoc Tukey's HSD tests ($p < .05$) revealed a similar pattern with less affiliation in the cross-race (AC – Eu) pairs than in the all-African Caribbean pairs (AC – AC). Thus, levels of affiliation were lowest in the cross-race pairs in both samples and highest in the same-race minority pairs.

Gender, race, and conversation style. Tables 1 – 4 reveal that many overall effects of speaker race were reflected in the use of assertion and affiliation among boys and girls separately. European boys used more assertion than South Asian boys, but there was no effect of speaker race comparing African Caribbean and European boys and no effect of speaker race on boys' use of affiliation in either sample. Conversely, although there were no effects of speaker race on girls' use of assertion in either sample, both South Asian and African Caribbean girls used significantly more affiliation than European girls. Thus, overall effects of speaker race on the use of assertion stem largely from differences in boys' conversations, whereas effects of speaker race on the use of affiliation stem from differences in girls' conversations.

Breaking down the use of conversation measures by gender reveals further variation in partner race effects. Boys used more assertive speech when talking with South Asian and African Caribbean partners than with European children. For boys, there were no partner race effects in the use of affiliation. Girls were more assertive toward a European girl than toward a South Asian girl, whereas they were more affiliative when speaking with an African Caribbean compared with a European conversation partner.

Although there was no overall interaction effect in use of assertion in the South Asian and European sample, the interaction effect was significant for the boys and the girls considered separately (see again Table 3). Both these interaction effects indicate that racial differences in the use of assertion were greater within cross-race pairs than between same-race pairs. However, closer inspection of the means in Table 2 reveals that whereas in same-race pairs, levels of assertion are broadly similar, in cross-race pairs, European boys are more assertive whereas European girls are less assertive than their South Asian partners. In the African Caribbean and European sample, European boys used more assertion in cross-race pairs, but there was no interaction effect relating to girls' use of assertion in same-race and cross-race pairs. There were interaction effects for boys' and girls' use of affiliation in both samples. Once again, racial differences in the use of affiliation were greater between children from different race groups in the same-race pairs than in the cross-race pairs.

A further analysis was conducted for each gender group that matched the comparison of assertion in cross-race pairs for the whole sample. Independent *t* tests revealed a significant difference only between the South Asian and European boys, the latter using a higher level of assertion in the cross-race pairs, $t(28) = 3.01$, $p < .01$, Cohen's $d = 1.14$.

Finally, we compared levels of affiliation across pair types for boys and girls in each sample separately to see if there were differences between same-race and cross-race pairs. There were no significant effects of pair type in the South Asian and European sample. In the African Caribbean and European sample, there was a significant effect of pair type for the girls' sample, $F(2, 105) = 6.87$, $p < .01$, $\eta^2 = .12$. Post hoc Tukey's HSD tests ($p < .05$) indicated that the same-race pairs differed significantly (Eu – Eu vs. AC – AC), with the higher levels of affiliation in the African Caribbean pair.

Playmate Choices

The outcome measure in terms of playmate choice was the racial group of the child that the pair selected (European, South Asian or African) or, when children could not agree on an answer together, unresolved. In the present study, two specific predictions were made about playmate choice. The fourth hypothesis predicted that in same-race pairs, children would tend to choose a partner from the racial in-group. The fifth hypothesis predicted that in cross-race pairs, there would be fewer agreements about the choice of a playmate. Figure 1 shows the proportion of selections for different playmates by racial group. The association between a pair's racial composition and choice of playmate was significant for the South Asian and European group, $\chi^2(4, N = 110) = 43.47$, $p < .001$, Cramer's V = .31, and for the African Caribbean and European group, $\chi^2(4, N = 104) = 109.49$, $p < .001$, Cramer's V = .51.

Same-race pairs. In the South Asian and European sample, most Eu – Eu pairs (73.8%) selected a European child as a possible playmate, and most SA – SA pairs (57.1%) selected a South Asian child as a possible playmate. Although both of these selections reflect in-group preferences, it is worth noting that the proportion of children in the SA – SA pairs who made this in-group selection was lower than in the same-race Eu – Eu pairs. In the African Caribbean and European sample, most Eu – Eu pairs (91.9%) selected a European child as a preferred playmate. AC – AC pairs tended to choose an African Caribbean child as a preferred playmate (65.5%), and again, these minority-group pairs showed less marked in-group preference than the European majority pairs.

Cross-race pairs. In the cross-race pairs (SA – Eu), there was a relatively high proportion of unresolved conversations (27.3%). Almost half (48.5%) of the cross-race pairs selected a European child as a preferred playmate and fewer (24.2%) cross-race pairs selected a South Asian child. Cross-race pairs (AC – Eu) were also less successful in reaching an agreement about their choice of playmate (i.e., 37.1% of the conversations were unresolved). In these cross-race pairs, a European playmate (42.9%) was chosen more often than an African Caribbean playmate (20%).

Gender, race, and playmate choices. There were no significant gender differences in terms of the race of the chosen playmate, both between and within the two different samples.

Race, Conversation, and Playmate Choices

We proposed three hypotheses relating to the links between race, conversation, and playmate choices. The sixth hypothesis predicted that there would be high levels of assertion and low levels of affiliation in

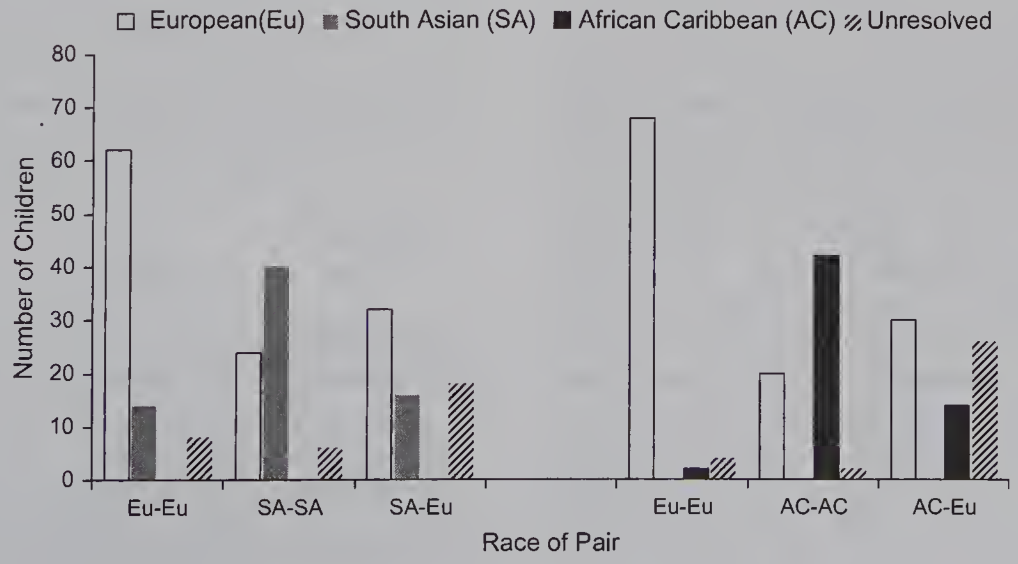


Figure 1. Children's choices of playmate by race and race of the pair.

Table 5

Mean Levels (and SDs) of Assertion and Affiliation Within a Pair by Outcome Playmate Preference for the South Asian and European Sample

	South Asian	European	Unresolved	*F*	*df*	η^2
All						
Assertion	4.96 (1.41)$_{a}$	5.36 (1.28)$_{ab}$	5.82 (0.91)$_{b}$	6.27**	2, 220	.06
Affiliation	4.30 (1.20)$_{a}$	4.13 (1.16)$_{a}$	2.31 (0.87)	31.25***	2, 220	.23
Boys only						
Assertion	4.67 (1.32)	5.51 (1.02)$_{a}$	5.94 (0.63)$_{a}$	7.66**	2, 110	.13
Affiliation	4.49 (1.08)$_{a}$	4.41 (1.15)$_{a}$	2.38 (0.72)	25.62***	2, 110	.32
Girls only						
Assertion	5.26 (1.46)$_{a}$	5.21 (1.48)$_{a}$	5.70 (1.14)$_{a}$	0.77	2, 110	.00
Affiliation	4.10 (1.30)$_{a}$	3.85 (1.11)$_{a}$	2.24 (1.01)	8.65**	2, 110	.15

Note. Means with the same subscript (within rows) are *not* significantly different from each other according to Tukey's honestly significant difference tests ($p < .05$).
$^{**}p < .01$. $^{***}p < .001$.

unresolved conversations. The seventh hypothesis predicted that same-race pairs who made an in-group choice of playmate would show high levels of affiliation. The eighth hypothesis predicted that in same-race pairs who made an out-group choice of playmate, there would be low levels of affiliation. Separate ANOVAs were conducted to examine the relations among pair composition, choice of playmate, and each of the conversation measures. Tables 5 and 6 present the mean levels of assertion and affiliation by conversation outcome (i.e., playmate choices), along with associated statistics, for the South Asian and European and African Caribbean and European samples, respectively. Tables 5 and 6 also show data and associated statistics broken down for each gender group.

Resolved and unresolved conversations. In both samples, variations in the use of assertion and affiliation were associated with the choices of playmate that pairs made. In the African Caribbean and European sample, levels of assertion were higher when conversations were unresolved than when both children agreed on a choice of playmate (although overall the effect was only marginally significant). In the South Asian and European sample, conversations that were unresolved were also associated with higher levels of assertion than when children agreed an answer. However, post hoc tests indicated that levels of assertion in unresolved conversations only differed significantly from conversations when children agreed on a South Asian playmate and not from when they agreed on a European playmate (see Table 5). In both samples, levels of affiliation were significantly lower in unresolved conversations than in conversations where children agreed a response together, which is consistent with the prediction made in the sixth hypothesis.

Affiliation and playmate choices. To test our seventh and eighth hypotheses, we examined the relation

Table 6

Mean Levels (and Standard Deviations) of Assertion and Affiliation Within a Pair by Outcome Playmate Preference for Children in the African Caribbean and European Sample

	African Caribbean	European	Unresolved	*F*	*df*	η^2
All						
Assertion	4.84 (1.28)$_{a}$	4.91 (1.20)$_{a}$	6.00 (0.75)	2.89†	2, 208	.03
Affiliation	4.55 (1.52)$_{a}$	4.04 (1.18)$_{a}$	2.37 (1.01)	13.11***	2, 208	.12
Boys only						
Assertion	4.65 (1.03)$_{a}$	5.08 (1.04)$_{a}$	6.30 (0.73)	5.49**	2, 100	.11
Affiliation	4.53 (1.57)$_{a}$	4.40 (1.21)$_{a}$	2.39 (1.10)	8.07**	2, 100	.15
Girls only						
Assertion	5.05 (1.49)$_{ab}$	4.76 (1.32)$_{b}$	5.76 (0.68)$_{a}$	1.52	2, 108	.00
Affiliation	4.58 (1.48)$_{a}$	3.71 (1.06)$_{b}$	2.36 (0.96)$_{c}$	9.20***	2, 108	.15

Note. Means with the same subscript (within rows) are *not* significantly different from each other according to Tukey's honestly significant difference tests ($p < .05$).
$^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

between playmate choice and levels of affiliation for same-race pairs only who agreed on a choice of playmate (i.e., resolved their conversations). Analysis of data from both samples indicated that there were no significant interactions among pair type, choice, and levels of affiliation.

Gender, race, conversation, and playmate choices. Results for boys and girls in each sample generally mirrored those for the sample overall. There were, however, some interesting variations. In the South Asian and European sample, boys used significantly less assertion when agreeing on a South Asian choice of playmate than when conversations were unresolved or children chose a European playmate. African Caribbean and European boys were most assertive when conversations were unresolved and least assertive when an African Caribbean or European playmate was selected. Although trends linking girls' use of assertion to conversation outcomes were similar across samples (i.e., higher levels of assertion in unresolved conversations than in resolved conversations), post hoc tests did not establish significant differences between groups.

Boys and girls in both samples used significantly less affiliation when conversations were unresolved than when they agreed on a choice of playmate together. Among girls in the African Caribbean and European sample, levels of affiliation were also significantly higher when a pair agreed to choose an African Caribbean playmate compared with when they chose a European playmate (see again Table 6).

Comparison of All-European Pairs From Both Samples

The existence of two separate samples (i.e., South Asian and European and African Caribbean and European) made it possible to compare the two sets of data for children in all-European pairs. This comparison constituted a check on the equivalence of the two samples. Separate independent t tests showed that there were no differences between the two samples on any of the dependent measures.

Discussion

The present study examined the influence of race on children's conversations about and choices of possible playmates. Three key research goals were (a) to identify the relations between race and children's conversational style, (b) to examine how race influenced children's playmate choices (i.e., conversation outcomes), and (c) to explore the relations among race, conversation style, and playmate choices. The findings indicate that race influences the style and outcomes of children's conversations about possible friends. Moreover, this influence is moderated by gender and varies according to the racial mix of the children involved in interaction.

Our initial set of hypotheses focused on the relation between the race and the style or social dynamics of children's conversations. The first hypothesis was confirmed. That is, a child's race influences the levels of assertion and affiliation the child uses in a conversation. Specifically, European children were more assertive than South Asian and African Caribbean children. South Asian children were more affiliative than European children, but there was no difference in levels of affiliation between European and African Caribbean children. These effects of speaker race indicate that there are different conversational styles that are linked to race among children at this age, perhaps as a result of different socialization experiences associated with membership of different race groups. Thus, it could be that European children learn a more assertive conversational style from interactions with parents, peers, and others (who will often be members of the racial in-group), whereas South Asian children learn to be more affiliative than children from the other racial groups that were studied here. Hughes et al. (2006) have noted that ethnic socialization can be one way in which many minority parents seek to raise socially competent children in a racialized society. Children's conversational styles may also be affected by observing others in wider society or in the children's own cross-race interactions that entail different contexts for socialization for children from different racial groups. The present study indicates that, by 7 years of age, children from different racial groups have learned to reproduce these different conversational styles in interaction with their peers.

We found no support for our second hypothesis that, as a consequence of status differences between children from majority and minority groups, European (i.e., majority group) children would be more assertive than African Caribbean and South Asian (i.e., minority group) children in cross-race interaction. A simple explanation may be that children did not perceive status differences between the racial groups. However, it seems unlikely that children at this age are unaware of status differences associated with race. By middle childhood, children recognize that minority ethnic groups are associated with different stereotypes and that there are differences in relative status between minority groups (Davis, Leman, & Barrett, 2007; Lam & Leman, 2003). Moreover, our results do not show that children fail to respond to the race of a conversation partner or to the racial mix of children involved in interaction. On the contrary, there were numerous variations in children's use of affiliation and assertion linked to race. All children were more assertive with a South Asian partner, and more affiliative with a South Asian or African Caribbean partner, than with a European child. In the South Asian and European sample, race differences in children's use of affiliation were greater comparing same-race with cross-race pairs. This finding suggests that children responded to the dynamics of in-group and out-group interaction. In the African Caribbean and European sample, race differences in assertion were more marked in same-race pairs, whereas race differences in affiliation were more marked comparing same-race pairs. Thus, children were sensitive to the race of their partner, and these effects were visible in conversation. However, it was not the case that this resulted in majority-group children being more assertive toward minority-group children in conversation.

The failure to find support for the second hypothesis suggests that, in terms of conversation at least, it is not straightforwardly the case that racial majority children will dominate or assert themselves over minority-group children. Additionally, although European children may have a more assertive conversational style generally, this assertion does not appear to be deployed to a greater or lesser extent when speaking with a minority-group partner. It should be noted, however, that in the present study, the participating children came almost exclusively from highly ethnically heterogeneous and low-income neighborhoods, and their schools celebrated this racial diversity. In another context (or at another point in development), levels of assertion in conversations between majority- and minority-group children might not be so balanced.

Although status differences do not appear to affect levels of assertion in cross-race interaction, a key tenet of work on intergroup dynamics is that individuals affiliate more with members of their in-group than with members of an out-group (Tajfel & Turner, 1986). We thus proposed a third hypothesis that predicted higher levels of affiliation in same-race pairs than in cross-race pairs. Our third hypothesis was partly supported. In both samples, children were more affiliative in minority same-race than in cross-race pairs. The mean levels of affiliation of same-race majority pairs were somewhere in between those of cross-race pairs and same-race minority pairs, but this trend did not reach significance and generally effect sizes were rather small.

Higher levels of affiliation in minority pairs may be a general phenomenon or a consequence of using playmate choice as the topic of discussion in the present study. In this respect, it is possible that minority children felt the race-related aspects of this task more intensely than majority children because of its connotations for real-life friendship choices and intergroup contact. Race may be a more salient feature of friendship choice to minority-group children at this age because of their more negative social experiences or because of a more hostile social environment (Brown & Bigler, 2005; Spencer, Cunningham, & Swanson, 1995). The present study opens the possibility that racial cleavage in friendships and playmate choices that has been observed in naturalistic studies from preschool onward (Finkelstein & Haskins, 1983) may be associated with particular interpersonal dynamics between children. Of course, a question remains whether these features of interaction are causes or consequences of this racial cleavage.

A further intriguing finding to emerge from the present study was that race affects conversations differently for boys and girls. European boys were more assertive than both their African Caribbean and South Asian contemporaries. African Caribbean and South Asian girls were more affiliative than European girls. One interpretation of these effects of speaker race is that these differences in conversational style stem from the distinct socialization experiences of boys and girls from different racial groups. European boys may learn to be more assertive than African Caribbean and South Asian boys, and African Caribbean and South Asian girls may learn to be more affiliative than European girls.

When cross-race pairs (i.e., African Caribbean and European pairs and South Asian and European pairs) and same-race pairs (i.e., all-African Caribbean, all-South Asian, and all-European pairs) were compared, there were also differences in the conversation style that were linked to the race of a child's conversation partner. In terms of assertion, the influence of race in boys' conversations divided along majority–minority racial lines. European boys (i.e., the majority group) were more assertive than African Caribbean and South Asian boys (i.e., the minority groups), but this difference was especially marked in cross-race interactions. In other words, the influence of race in boys' conversations led to more asymmetry in levels of assertion when boys talked with boys from a different race group. This finding supports the observation that preschool boys' interactions are more dominance oriented than girls' interactions (Leaper & Smith, 2004) and that these differences continue into adulthood (Sidanius, Levin, Liu, & Pratto, 2000). It may be that, at 7 years of age, race constitutes a source of social status or asymmetry that fits easily into the dominance-oriented dynamics of boys' conversations.

In contrast to the boys, girls from all racial groups were more assertive toward a European conversation partner. In South Asian and European pairs, the European girls displayed very low levels of assertion. Compared with boys', assertion is less of a feature of girls' interactions (Leaper & Smith, 2004), and so may be considered a more negative characteristic in girls' discussions. By avoiding assertion, European girls may be positioning themselves (or are being positioned) closer to a gender norm in conversation, that is, inviting agreement rather than demanding it (see Leman et al., 2005). South Asian and African Caribbean girls were more affiliative speakers, and all the girls were more affiliative with an African Caribbean girl, which corresponds to the finding that children tend to give higher companionship and friendliness ratings to African Caribbean girls than to other race and gender groups (Aboud et al., 2003). The differences between the races were greatest when comparing levels of affiliation across same-race pairs than when comparing within cross-race pairs.

Our second set of hypotheses concerned children's choices of playmate. The fourth hypothesis, which predicted that same-race pairs would be more likely to choose a playmate from their racial in-group, was confirmed for all three racial groups. In some senses, this finding should not be surprising given that playmate preference delineates along racial lines from a very young age (see Finkelstein & Haskins, 1983); however, in-group preference was notably less pronounced in same-race minority pairs (i.e., African Caribbean or South Asian) than in same-race majority pairs (i.e., European). This fits with the results of some recent studies (Margie, Killen, Sinno, & McGlothlin, 2005; McGlothlin et al., 2005) indicating that U.S. minority-group children are more likely than majority-group children to think peers from different racial groups can be friends. As predicted by the fifth hypothesis, cross-race pairs had more difficulty reaching an agreement about the choice of a playmate than same-race pairs. This was probably because of conflicting in-group choices. When children in cross-race pairs agreed on a choice, most pairs opted for a European child as their preferred playmate.

These results are reminiscent of previous research (e.g., Aboud, 1988; Spencer & Markstrom-Adams, 1990) that suggested that minority-group children exhibit less preference for their own racial group compared to majority-group children. Our findings also fit with previous work that used attitude measures outside the context of race and identified more in-group bias in high- (than low-) status groups of children between 7 and 12 years of age (Bigler, Brown, & Markell, 2001). In this case, perhaps racial minority children desire a playmate from a higher status group (and perceive Europeans to be higher status). Our conversational paradigm is a novel means of eliciting information about children's racial understanding that can inform studies of racial attitudes and identification.

Our last set of hypotheses explored the links among race, conversations, and playmate choices. The sixth hypothesis was confirmed and, as predicted, there were higher levels of assertion and lower levels of affiliation in unresolved conversations. This finding underscores the predicted role of assertion and affiliation in decision-making processes. Children who are unable to agree will simply reassert their preference and continued disagreement leads children to disaffiliate.

Analysis of the data did not, however, confirm the seventh and eighth hypotheses. These predicted, respectively, that there would be high levels of affiliation when same-race pairs agreed on an in-group playmate and low levels of affiliation when same-race pairs agreed on an out-group playmate. At first glance, this finding suggests that our attempt to identify a link between conversation, intergroup dynamics, and decision making has failed. Indeed, it may be that while the race of a pair influences which playmate the pair chooses, levels of affiliation in conversation do not correlate with this decision. However, it should be noted that very few children in same-race pairs chose an out-group playmate, and a larger sample size may have revealed some significant effects. Moreover, in-group and out-group biases in adults may, sometimes unknowingly, affect decision-making processes and friendship choices (Mackie & Ahn, 1998). In the present study, given how lengthy children's discussion were, it seems surprising that intergroup dynamics did not affect affiliation and, in turn, children's friendships choices. Further work is needed to establish the ways in which race, conversation, and friendship choices interrelate because it is through conversation and discussion that meaningful contact between children happens.

Some caution should be applied when extending these findings to *all* interactions and playmate choices. In the present study, children were drawn from a school with a heterogeneous mix of races. Although this mix mirrors many contemporary urban environments, the findings from this study may be different in other sociocultural contexts and with other racial groups. Moreover, other factors may influence conversation dynamics between children. For example, Brown and Bigler (2002) found several different effects of minority status on intergroup attitudes that were related to the size, salience, and status of the groups as well as children's ages and the measures employed. Our findings and those of other studies suggest that these and other factors may also relate to variations in conversation dynamics. Indeed, the context in which conversations happen provides not only a backdrop to those discussions but also has the potential to act as a further influence on the dynamics of those conversations. From a practical perspective, interventions seeking to reduce racial prejudice need to be mindful of these broader influences on attitudes and behavior, as well as the specifics of conversations that were explored here.

In terms of intergroup relations, the quality of conversation in the cross-race pairs is of particular interest because contact between children of different races has been suggested as a way to improve intergroup relations, reduce prejudice, and foster positive intergroup attitudes (Pettigrew & Tropp, 2006). In our cross-race pairs, children had some difficulty agreeing on a choice of playmate, and when they did agree, they more often settled on a European (majority group) child rather than a South Asian or African Caribbean (minority group) child. Our findings suggest that interventions focusing on contact between children from different races (see Cameron et al., 2006) need to focus not only on achieving contact but also on guiding that contact in ways that challenge children's existing friendship preferences to ensure sustained interaction between children from different groups. Particular attention needs to be paid to fostering affiliative interactions in cross-race pairs. In this respect, our friendship task certainly did not encourage affiliation, given that the task pitted children's in-group preferences against one another. Other, racially neutral tasks—for instance, free play, problem solving, or collaborative learning tasks—may provide better opportunities for children from different racial groups to affiliate. Such opportunities are needed to address the racial cleavage in friendship and playmate choices in place early in childhood and continuing through to adulthood (Hallinan, 1982). The present study demonstrates not only that race affects children's choices of playmate at 7 years but also that race is a powerful force that affects the style and dynamics of children's conversations.

References

Aboud, F. E. (1988). *Children and prejudice*. Cambridge, MA: Blackwell.

Aboud, F. E., & Doyle, A. B. (1996). Does talk of race foster prejudice or tolerance in children? *Canadian Journal of Behavioral Science*, *28*, 161–170.

Aboud, F. E., & Fenwick, V. (1999). Exploring and evaluating school-based interventions to reduce prejudice. *Journal of Social Issues*, *55*, 767–786.

Aboud, F. E., & Levy, S. R. (2000). Interventions to reduce prejudice and discrimination in children and adolescents. In S. Oskamp (Ed.), *Reducing prejudice and discrimination* (pp. 269–293). Mahwah, NJ: Erlbaum.

Aboud, F. E., Mendelson, M. J., & Purdy, P. (2003). Cross-race peer relationships and friendship quality. *International Journal of Behavioral Development*, *27*, 165–173.

Abrams, D., Rutland, A., Cameron, L., & Marques, J. M. (2003). The development of subjective group dynamics: When in-group bias gets specific. *British Journal of Developmental Psychology*, *21*, 155–176.

Bigler, R. S., Brown, C. S., & Markell, M. (2001). When groups are not created equal: Effects of group status on the formation of intergroup attitudes in children. *Child Development*, *72*, 1151–1162.

Boulton, M., & Smith, P. K. (1992). Racial preferences and perceptions among Asian and European British middle school children. *Social Development*, *1*, 55–65.

Brown, C. S., & Bigler, R. S. (2002). Effects of minority status in the classroom on children's intergroup attitudes. *Journal of Experimental Child Psychology*, *83*, 77–110.

Brown, C. S., & Bigler, R. S. (2005). Children's perceptions of discrimination: A developmental model. *Child Development*, *76*, 533–553.

Cameron, L., Rutland, A., Brown, R. J., & Douch, R. (2006). Changing children's intergroup attitudes towards refugees: Testing different models of extended contact. *Child Development*, *77*, 1208–1219.

Carli, L. L. (1989a). Gender differences in interaction style and influence. *Journal of Personality and Social Psychology*, *56*, 565–576.

Carli, L. L. (1989b). Gender differences in interaction style and influence: Correction. *Journal of Personality and Social Psychology*, *57*, 964.

Davis, S. C., Leman, P. J., & Barrett, M. (2007). Children's implicit and explicit ethnic group attitudes, ethnic group identification, and self esteem. *International Journal of Behavioral Development*, *31*, 514–525.

Doyle, A. B., & Aboud, F. E. (1995). A longitudinal study of white children's racial prejudice as a social-cognitive development. *Merrill-Palmer Quarterly*, *41*, 209–228.

Finkelstein, N. W., & Haskins, R. (1983). Kindergarten children prefer same color peers. *Child Development*, *54*, 502–508.

Fishbein, H. D., & Imai, S. (1993). Pre-schoolers select playmates on the basis of gender and race. *Applied Developmental Psychology*, *14*, 303–316.

Garcia-Marques, T., & Mackie, D. M. (2001). The feeling of familiarity as a regulator of persuasive processing. *Social Cognition*, *19*, 9–34.

Gonzalez, R., & Griffin, D. (2000). The statistics of interdependence: Treating dyadic data with respect. In W. Ickes & S. W. Duck (Eds.), *The social psychology of personal relationships* (pp. 181–213). Chichester, UK: Wiley.

Hallinan, M. T. (1982). Classroom racial composition and children's friendships. *Social Forces, 61*, 56–72.

Hallinan, M. T., & Teixeira, R. A. (1987a). Opportunities and constraints: Black-white differences in the formation of interracial friendships. *Child Development, 58*, 1358–1371.

Hallinan, M. T., & Teixeira, R. A. (1987b). Students' interracial friendships: Individual characteristics, structural effects and racial differences. *American Journal of Education, 95*, 563–583.

Howes, C., & Wu, F. (1990). Peer interactions and friendships in an ethnically diverse school setting. *Child Development, 61*, 537–541.

Hughes, D., Rodriguez, J., Smith, E. P., Johnson, D. J., Stevenson, H. C., & Spicer, P. (2006). Parents' ethnic-racial socialization practices: A review of research and directions for future study. *Developmental Psychology, 42*, 747–770.

James, D., & Drakich, J. (1993). Understanding gender differences in amount of talk: A critical review of the research. In D. Tannen (Ed.), *Gender and conversational interaction* (pp. 281–306). New York: Oxford University Press.

Kenworthy, J. K., Turner, R. N., Hewstone, M., & Voci, A. (2005). Intergroup contact: When does it work and why? In J. Dovidio, P. Glick, & L. Rudman (Eds.), *Reflecting on the nature of prejudice* (pp. 278–292). Cambridge, MA: Blackwell.

Killen, M., & Stangor, C. (2001). Children's reasoning about inclusion and exclusion in gender and race peer group contexts. *Child Development, 72*, 174–186.

Kraemer, H. C., & Jacklin, C. N. (1979). Statistical analysis of dyadic social behavior. *Psychological Bulletin, 86*, 217–224.

Lam, V. L., & Leman, P. J. (2003). The influence of gender and ethnicity on children's inferences about toy choice. *Social Development, 12*, 269–287.

Leaper, C. (1991). Influence and involvement in children's discourse: Age, gender, and partner effects. *Child Development, 62*, 797–811.

Leaper, C. (2000). Gender, affiliation, assertion, and the interactive context of parent-child play. *Developmental Psychology, 36*, 381–393.

Leaper, C., & Smith, T. E. (2004). A meta-analytic review of gender variations in children's language use: Talkativeness, affiliative speech, and assertive speech. *Developmental Psychology, 40*, 993–1027.

Leaper, C., Tenenbaum, H. R., & Shaffer, T. G. (1999). Gender effects on the communication strategies of African American children from low-income, urban backgrounds. *Child Development, 70*, 1489–1503.

Lease, A. M., & Blake, J. J. (2005). A comparison of majority-race children with and without a minority-race friend. *Social Development, 14*, 20–41.

Leman, P. J., Ahmed, S., & Ozarow, L. (2005). Gender, gender relations and the social dynamics of children's conversations. *Developmental Psychology, 41*, 64–74.

Mackie, D. M., & Ahn, M. N. (1998). Ingroup and outgroup inferences: When ingroup bias overwhelms outcome bias. *European Journal of Social Psychology, 28*, 343–360.

Maltz, D., & Borker, R. (1982). A cultural approach to male-female mis-communication. In J. Gumperz (Ed.), *Language and social identities* (pp. 196–216). Cambridge, UK: Cambridge University Press.

Margie, N. G., Killen, M., Sinno, S., & McGlothlin, H. (2005). Minority children's intergroup attitudes about peer relationships. *British Journal of Developmental Psychology, 23*, 251–269.

McGlothlin, H., & Killen, M. (2006). Intergroup attitudes of European American children attending ethnically homogeneous schools. *Child Development, 77*, 1375–1386.

McGlothlin, H., Killen, M., & Edmonds, C. (2005). European American children's intergroup attitudes about peer relationships. *British Journal of Developmental Psychology, 23*, 227–249.

Pettigrew, T. F. (1997). Generalized intrergroup contact effects on prejudice. *Personality and Social Psychology Bulletin, 23*, 173–185.

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Interpersonal Relations and Group Processes, 90*, 751–783.

Powlishta, K., Serbin, L. A., Doyle, A., & White, D. R. (1994). Gender, ethnic, and body-type biases: The generality of prejudices in childhood. *Developmental Psychology, 30*, 526–536.

Quintana, S. M., Aboud, F. E., Chao, R. K., Contreras-Grau, J., Cross, W. E., Jr., Hudley, C., et al. (2006). Race, ethnicity, and culture in child development: Research and future directions. *Child Development, 77*, 1129–1141.

Sidanius, J., Levin, S., Liu, J., & Pratto, F. (2000). Social dominance orientation, anti-egalitarianism and the political psychology of gender. *European Journal of Social Psychology, 30*, 41–67.

Spencer, M. B., Cunningham, M., & Swanson, D. P. (1995). Identity as coping: Adolescent African-American males' adaptive responses to high risk environments. In H. W. Harris, H. C. Blue, & E. H. Griffith (Eds.), *Racial and ethnic identity* (pp. 31–52). New York: Routledge.

Spencer, M. B., & Markstrom-Adams, C. M. (1990). Identity processes among racial and racial minority children in America. *Child Development, 61*, 290–310.

Tajfel, H., & Turner, J. C. (1986). The social identity theory of intergroup behavior. In S. Worchel & L. W. Austin (Eds.), *Psychology of intergroup relations* (pp. 7–24). Chicago: Nelson-Hall.

Tropp, L. R., & Pettigrew, T. F. (2005). Relations between intergroup contact and prejudice among minority and majority status groups. *Psychological Science, 16*, 951–957.

United Kingdom Government, Office for National Statistics. (2005). *Census 2001*. Retrieved August 10, 2005, from http://www.statistics.gov.uk/cci/nugget.asp?id=273

Wright, S. C., & Tropp, L. R. (2005). Language and intergroup contact: Investigating the impact of bilingual instruction on children's intergroup attitudes. *Group Processes and Intergroup Relations, 8*, 309–328.

Child Development, September/October 2008, Volume 79, Number 5, Pages 1344–1356

Where Is the Essence? Developmental Shifts in Children's Beliefs About Internal Features

George E. Newman and Frank C. Keil
Yale University

The present studies investigated children's and adults' intuitive beliefs about the physical nature of essences. Adults and children (ranging in age from 6 to 10 years old) were asked to reason about 2 different ways of determining an unknown object's category: taking a tiny internal sample from any part of the object (distributed view of essence) or taking a sample from one specific region (localized view of essence). Results from 3 studies indicated that adults strongly endorsed the distributed view, and children showed a developmental shift from a localized to distributed view with increasing age. These results suggest that even children go beyond mere placeholder notions of essence, committing to conceptual frameworks of how essences might be physically instantiated.

Adults and children alike seem to believe that many sorts of things have essences. In its everyday sense, an essence is often thought of as something intrinsic to an entity that causes that entity to be the kind of thing that it is. For example, the essence of a tiger would be some unique characteristic, such as DNA, that is shared by all tigers and causes tigers to have their distinct tiger properties (Sober, 1994). Such beliefs are part of a cognitive bias known as "psychological essentialism" (Medin & Ortony, 1989), a bias that not only embraces the notion of an essence but also the idea that this deep, underlying nature is somehow causally responsible for an entity's category membership and its phenomenal properties (Gelman, 2003).

Much of the support for psychological essentialism comes from studies with young children. For example, children as young as 4 years old appear to understand that internal causes are more to likely determine an objects' behavior and appearance than are external ones (Gelman, 2003; Gottfried & Gelman, 2005; Inagaki & Hatano, 2002; Sobel, Yoachim, Gopnik, Meltzoff, & Blumenthal, 2007), and they are likely to view internal properties as vital to determining an animal's category membership (Gelman & Wellman, 1991; Keil, 1989). Young children will also extend internal, nonvisible features to novel category instances (e.g., Gelman & Coley, 1990; Gelman & Markman, 1987). For example, if children are taught a property about leaf insects, they are likely to generalize that property to similarly named organisms (other insects) instead of similar-looking objects (other leaves), even when surface appearance conflicts with category membership (Gelman & Markman, 1987).

The emergence of such reasoning in young children underscores the strength of essentialist commitments. Yet, the ease with which children seem to reference internal, nonvisible features presents somewhat of a dilemma. As adults, we are at least familiar with the concepts of cells, molecules, and DNA. Thus, there is at least some basis for essentialist beliefs to be grounded in real, physical structures—even if those connections to reality are often inaccurate (Hull, 1965; Keil & Richardson, 1999; Mayr, 1982). But for children as young as 4 or 5, who presumably know nothing about such structures, it is unclear what type of representation they are using to guide their inferences about these unseen, internal properties. In other words, in the absence of concrete theories about actual biological mechanisms, what does it mean to take an essentialist stance?

Some researchers have suggested that essences are understood more as causal "placeholders," rather than as actual physical entities (Gelman, 2003; Medin & Ortony, 1989). Thus, one could believe that there is some underlying property that unites category members, without knowing exactly what that property actually is or where it might be located. It might be that for young children—and for that matter, many adults—psychological essentialism consists of nothing more than the belief that certain kinds of entities,

---

This research was supported by a National Institutes of Health Grant RD37-HD023922 to Frank Keil. We thank the staff, parents, and students at Robert Healy Elementary School, The Barstow School, Oak Hill Elementary School, and Walter Derynoski Elementary. Special thanks to Patricia Balbas for testing the children, and to Woo-kyoung Ahn, Lance Rips, and two anonymous reviewers for their helpful comments on earlier versions of the manuscript.

Correspondence concerning this article should be addressed to George E. Newman, Department of Psychology, Yale University, New Haven, CT 06520-8205. Electronic mail may be sent to george.newman@yale.edu.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7905-0009

such as animals, have essences, and that those essences are somehow causally responsible for each member of a kind sharing certain properties. In this "barest placeholder" sense, the child is completely agnostic about particular causal relationships or mechanism and believes nothing beyond the very general idea that essences cause surface properties in ways that remain completely mysterious (see Strevens, 2000, for a more detailed account of different notions of psychological essentialism).

An alternative to the barest placeholder account would argue that even though children clearly do not have knowledge about specific biological mechanism such as genes and gene regulation, they still might possess much more abstract hunches about likely causal pathways and patterns through which essences have their effects. There is ample precedent for the idea that young children, even preschoolers, can have abstract expectations about causal patterns in a domain while not having any idea about specific mechanism (Inagaki & Hatano, 2002; Kalish & Gelman, 1992; Simons & Keil, 1995). For example, children expect animals and machines to have different insides even when they lack specific predictions about what those insides should look like (Simons & Keil, 1995). By this account, even if young children obviously do not know anything about DNA or the molecular basis of gene expression, they still might have some skeletal hunches about the ways in which essences are physically manifested.

Moreover, there is evidence to suggest that children may often conceive of essences as "very tiny, internally located" structures. For example, Springer and Keil (1991) found that if young children are asked to describe how various types of properties might be transmitted from parents to offspring, they are likely to endorse "Gemmulic" theories of biological transmission. In these studies, children reported, for example, that when a brown mother dog has a puppy, the puppy is likely to be brown because, "some very tiny brown pieces went from the mother to the puppy, . . . got into the puppy's skin, and the puppy turned brown." Springer and Keil derived the name "gemmules" from early theories in biology that proposed that the fetus inherits tiny pieces or copies of organs emanating from its mother's body (Appleman, 1979; Mayr, 1982).

The aim of the present studies was to investigate whether children (and adults) may go beyond the barest placeholder sense of essence, to make auxiliary assumptions about the way in which essences are physically realized within an entity. For obvious reasons, it is difficult to directly ask children and adults about the physical nature of essences because people rarely refer to the concept of "essence" per se. Instead, similar to previous work (e.g., Keil, 1989), the present studies used "discovery" scenarios in which the category of an object is unknown, and participants must decide the best way to identify that object's category. This method assumes that beliefs about the category-identifying feature (a feature that is specific to one particular kind and no other) should map onto beliefs about the essential feature (a feature that causes one kind to be different from all others).

This method is about identifying categories, not individuals. With individuals, idiosyncratic surface features are often paramount. A person's tattoo, fingerprint, or dental work may be all one needs to confidently identify a person, even though those marks are not at all considered central to the individual. With categories, however, it is far less common to assume that there is some sort of causally irrelevant "accidental" marker of category membership. To be sure, surface traits can indicate category membership, but we assume they work because they are lawful products of internal causes. Hence, the focus on categories in these studies avoids some concerns about idiosyncratic surface properties that might arise for individuals. In addition, by focusing exclusively on internal parts, we are selecting just those properties that prior work has argued as closest to beliefs about internal determinants of kind (e.g., Gelman, 2003; Keil, 1989).

So, what sorts of assumptions might children make about the physical nature of essences? Adults, for instance, tend to believe there are kind-specific microscopic structures (DNA and molecules) that are common to members of naturally occurring categories. For living kinds, such beliefs may actually be inaccurate as the actual genetic basis of a species is always a probabilistic distribution of highly similar but not identical genetic sequences (Hull, 1965; Keil & Richardson, 1999; Mayr, 1982). Moreover, adults often appreciate that the same microstructure is homogeneously distributed throughout an organism—for instance, that the DNA in my finger is the same as the DNA in my stomach—even though they may not fully appreciate how microstructure in these different locations is responsible for morphological differences. One way to characterize such beliefs is that they reflect a "distributed" notion of essential features—that is, that members of a category share a common feature, which is distributed throughout the inside of each category member. Thus, despite the existence of different internal structures within a particular entity (e.g., the heart, the brain, or the lungs), the same essential feature (DNA) is present throughout. In terms of categorization, because the critical feature is distributed, samples from any internal location should be capable of identifying its kind; a sample from one part should be as equally identifying of category membership as a sample from another part.

Assuming that adults' beliefs about the essential features of natural kinds are best characterized as *distributed*—a possibility we explored in the first study—a critical question is, how do beliefs about distributed essence develop? Do children constrain their inferences about essential features in a manner similar to adults?

It may be that for natural kinds, children instead tend to think of the essential feature as a single entity that is located in one special place. This "localist" notion of essential features might be analogous to the idea that there is a "seat of soul" such as the pineal (Bloom, 2004; Descartes, 1985), a highly focal point out of which causal effects emanate. Thus, rather than thinking of essential features as distributed throughout an entity, children may think that there is a single focal place within an object that determines category membership. This view might predict that, unlike the distributed account, there could be only one location where one could look to identify the category of an object. In all the three studies, we contrasted the distributed view against the localized view to determine how children of different ages and adults may think about the physical nature of essences.

If children do make some assumptions about how identifying features are physically realized within an object, a second critical question asks about the domain specificity of those notions. Consider first the contrast between biological kinds and other nonliving natural kinds. There is broad consensus that both are thought to possess essential features (e.g., Ahn et al., 2001; Medin & Ortony, 1989; Putnam, 1975; but also see Malt, 1994; Strevens, 2000 for alternative accounts); but the nature of those features is quite different. Cut a piece of gold into any number of pieces and all pieces are "essentially" the same. Cut a dog into small pieces (only as a thought experiment, please), and it is still clear that most pieces will be very different from each other. Indeed, most nonbiological kinds (e.g., elements such as gold and sulfur, compounds such as water and jade, and mixtures such as dirt and sand) are normally considered to have homogeneous internal structures. Such things are normally substances that take mass nouns—although, not necessarily (e.g., an icicle or a snowflake). In comparison, living kinds are predominately bounded entities with heterogeneous internal structures and take count nouns; only with a layer of more sophisticated biological knowledge, do the deeper commonalities in all biological tissue in an organism become apparent.

Because of this major contrast and because very young children do have a clear sense of count/mass distinctions as well as a some notions about substances (e.g., Au, Romo, & DeWitt, 1999; Dickinson, 1987; Stavy, 1991), our first two studies contrast living and nonliving natural kinds. Because even a modest amount of experience with broken up pieces of rocks or minerals would reveal that nonliving natural kinds are homogeneous, a shift to a distributed view of essences might occur earlier in development for homogenous substances than for living kinds and in fact might be a basis for later emerging analogous notions about living kinds. Using the method outlined above, our first two studies compare beliefs about essential features for living kinds with beliefs about essential features for nonliving natural kinds.

A second major contrast is between all natural kinds and artifacts. It is controversial whether artifacts should even be thought of having essential features anything like those of natural kinds. For example, they have been argued to either have no essences (Sloman & Malt, 2003) or to have essences in the form of the intended category of their creators (Bloom, 2004). Moreover, given sharply divergent intuitions among elementary school children about the relevance of surface properties to category membership of artifacts versus natural kinds (Gelman, 1988; Keil, 1989), the artifact/natural kind divide may reveal contrasting intuitions about distributed versus localized accounts. Our third study explores this contrast while also using a converging method to explore the generality of the first two studies. In particular, we predict that it would be implausible at any age to have a distributed view of artifacts (other than for synthesized substances such as plastics).

In short, these studies use intuitions about features needed for categorization (identifying features) as a way of asking whether young children merely believe in essential features that somehow determine category membership in a completely unspecified manner or whether they also embed their sense of essential features in a more constraining framework that specifies how those features are physically realized within objects. We predict that even young children will show judgments that go beyond barest placeholder notions. We further predict that these intuitions will vary across broad domains such as substances, living kinds, and artifacts, with distributed views emerging first for substances, then for living kinds, and never for artifacts.

Experiment 1

The purpose of Experiment 1 was to determine whether the distinction between the distributed versus the localist accounts was a psychologically useful one, and second, whether adults show any strong endorsement of either view.

Method

Participants. Thirty-one undergraduates participated in exchange for course credit.

Stimulus materials and procedure. In a within-subjects design, participants were given questionnaires that described a situation in which scientists discovered objects whose category was unknown. They read, for example:

> Many years ago, a giant glacier moved across the land and froze some things inside. A group of scientists discover the glacier and they notice that a large animal is frozen inside. They want to know what kind of animal it is. The scientists decide that they will take really tiny pieces from the inside of the animal. Each piece will be about the size of a pea, and then the scientists will use all their tools and technology to look at these tiny pieces, and they will be able to see everything that is inside each piece.

Participants were then asked to rate the degree to which they agreed with the each of the following statements:

1. "It doesn't matter where they take the pieces from, any piece will be able to tell them what kind of animal it is" (distributed statement).
2. "There is only one special piece that will tell them what kind of animal it is" (localized statement).

For four different natural objects (a large animal, a small animal, a rock, and a piece of metal), participants rated on a 0–9 scale (where higher numbers denoted greater agreement) the degree to which they agreed with the distributed and localized statements. The wording was similar for each item except that the words "small animal," "rock," and "piece of metal" were substituted as appropriate. We created four different versions of the survey that varied the order in which the different scenarios were presented as well as whether the localized or distributed statement appeared first.

Results

The results from Experiment 1 are reported in Table 1. Adult participants were overwhelmingly more likely to agree with the distributed approach, as confirmed by a mixed-model analysis of variance (ANOVA) with statement type (distributed vs. localized) and object type (animal vs. substance) as within-subjects factors. This analysis revealed a significant main effect of statement type on participants' ratings, $F(1, 30) = 128.69$, $p < .001$, partial $\eta^2 = .811$; as predicted, participants were more likely to agree with the distributed statement ($M = 6.93$, $SE = .27$) than with the localized statement ($M = 1.34$, $SE = .26$). We also observed a significant main effect of object type, $F(1, 30) = 10.78$, $p = .003$, partial $\eta^2 = .264$; participants gave higher ratings to substances ($M = 4.41$, $SE = .09$) than to animals ($M = 3.85$, $SE = .17$). We also found a significant interaction between statement type and object type, $F(1, 30) = 6.98$, $p < .05$, partial $\eta^2 = .188$. This interaction was driven by differences in ratings of the distributed statement. Participants were more likely to agree with the distributed statement for substances ($M = 7.59$, $SE = .27$) than for animals ($M = 6.27$, $SE = .40$), $t(30) = 3.22$, $p < .01$. Agreement with the localized statements, however, was not significantly different between items ($p = .62$). We ran an additional ANOVA that included which survey version participants completed (i.e., presentation order) as a between-subjects factor. There were no main effects or interactions of presentation order on participants' ratings ($F < 1$).

Table 1
Mean Agreement (Standard Deviation) With Localized and Distributed Statements for Each Item in Experiment 1

Item	Localized	Distributed
Large animal	1.40 (1.61)	6.40 (2.25)
Small animal	1.47 (1.79)	6.15 (2.57)
Rock	1.26 (1.65)	7.19 (2.01)
Metal	1.21 (1.67)	7.98 (1.36)

Note. Judgments were made on a 0–9 scale, where higher numbers denoted greater agreement.

Agreement with distributed statements was negatively correlated with agreement with localized statements (for substance items, $r = -.79$, $p < .001$; for animal items, $r = -.52$, $p < .01$).

Discussion

The results indicated that there is strong agreement among adults about the physical nature of essential features. Adults appear to think of such features as distributed throughout the inside of natural kinds and not localized to one particular region or part.

We also observed a significant interaction between statement type and object type. Adults were more likely to agree with distributed statements that referred to substances than distributed statements that referred to animals. One interpretation of this result is that substances are simply known to be more internally homogenous than animals, and the interaction with object type reflected this difference. Another interpretation might be that although adults overwhelmingly endorse distributed views for both animals and substances, they could still think of the essential features of animals as more localized. In other words, though adults have explicitly learned that the internal causal property in animals (DNA) is distributed, perhaps more developmentally primitive ways of viewing these structures (i.e., localist notions) still compete with this view. Therefore, data from children (Experiments 2 and 3) may shed light on this issue, as they provide an important comparison not only for when distributed notions develop but also what types of knowledge might be used to support such beliefs. If, for example, children think of substances as distributed before animals, it may because they use their knowledge about homogeneous substances to support more complicated inferences about the underlying essence of living kinds.

Experiment 2

The purpose of Experiment 2 was to explore when distributed views of essential features develop and any potential differences between substances and animals. We adapted the adult procedure from Experiment 1 to be suitable for elementary school children. We tested elementary school children as a way of dissociating the degree to which notions of distributed microstructure might stem from explicit knowledge about actual microstructure (DNA and molecules) versus more naïve assumptions about the nature of natural kinds. Presumably, all our adult participants (college undergraduates) had been exposed to the concepts of DNA and molecules in some formal sense. However, children are typically not exposed to these concepts in school until sixth- and seventh-grade biology and chemistry, and it is not typically until high school that children learn about the periodic table or biological processes such as cell division, genetic recombination, and so forth. This of course does not rule out the possibility of our participants' learning about DNA from other sources, such as parents, older siblings, or television crime shows. It does, however, limit the degree to which children have formally learned, for instance, that DNA consists of chromosomes and base pairs, which are replicated in every cell in the body. Thus, it is fair to say that children and our sample of college adults were substantially different in terms of their explicit knowledge about actual microscopic bodies.

Method

Participants. Participants were 67 children, including 20 kindergarteners (*M* age = 6.0, range = 5.6–6.7), 23 second graders (*M* age = 8.1, range = 7.7–8.6 months), and 24 fourth graders (*M* age = 10.2, range = 9.7–10.7 months), each of whom was interviewed individually. An additional 9 students were tested but did not complete the session. Altogether there were 36 boys and 31 girls, roughly evenly distributed across grade. Participants were from middle- and upper-middle-class backgrounds and were recruited at public and private elementary schools. Ethnicity breakdowns were as follows: Asian (*n* = 41), Caucasian (*n* = 21), Hispanic (*n* = 4), and African American (*n* = 1). A polling of the teachers confirmed that children were not receiving instruction about DNA or chemical elements in class.

Stimulus materials and procedure. Participants were first shown a picture that depicted a large block of ice with objects frozen inside. The objects were created to loosely resemble four different natural objects (identical to Experiment 1). The experimenter said:

> Many years ago a giant block of ice moved across the land and froze some things inside. A group of scientists discover the block of ice and they want to know what kinds of things are frozen inside.

Participants were then presented with a card that "zoomed in" on one of the objects. The experimenter then said:

> Scientists discover that a large animal is inside the block of ice and they want to know what kind of animal it is. To figure it out, the scientists decide that they will take really tiny pieces from the inside of the animal. Each piece is really, really, tiny. It's about the size of a pea. The scientists will use all their tools and technology to look at these tiny pieces, and they can see everything that is inside each piece.

Participants were then shown two pictures of scientists thinking about the tiny pieces (see Figure 1). The experimenter said:

> There is a disagreement between the two scientists. One scientist thinks that it doesn't matter which piece they look at to figure out what kind of animal it is, because any of the pieces they look at can tell

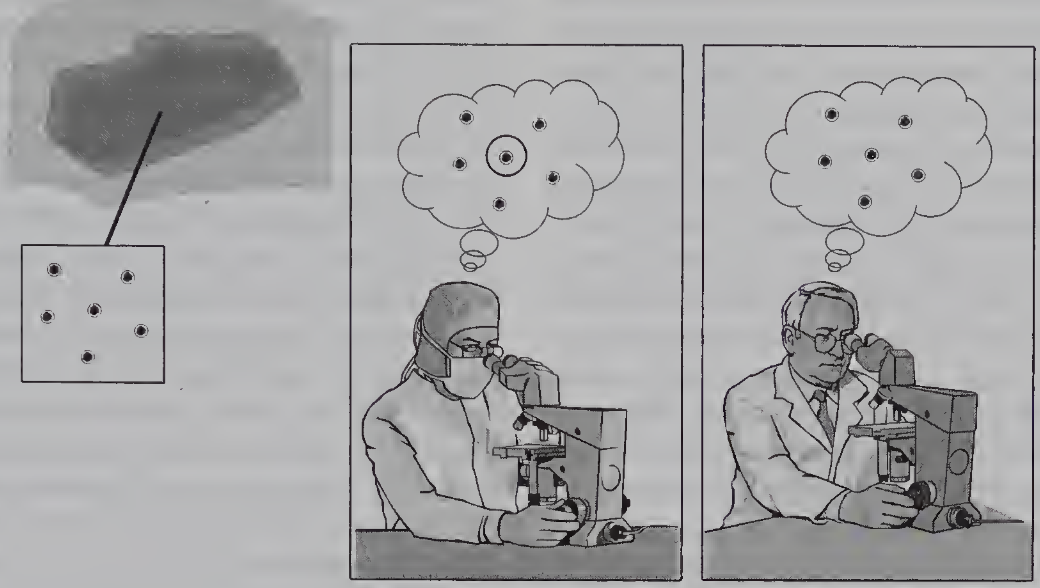


Figure 1. Sample pictures used in Experiment 2.

> them. But the other scientist thinks that there is only one piece that can tell them about what kind of animal it is, and they must find that one piece to figure it out.

Participants were asked to decide which scientist they thought was correct. After they responded, either verbally or by pointing, they were also asked to explain why they chose that answer. This process was repeated for the remaining three items (a small animal, a rock, and a piece of metal). The protocol was similar for each item, except that the words "small animal," "rock," and "piece of metal" were substituted as appropriate. The order in which the statements were presented was balanced within each participant so that for half of the items, a child was given the distributed statement first (any piece can tell them), whereas for the other half of the items, the localized statement (only one piece can tell them) was presented first. Additionally, two sets of distributed and localized images were created (where we crossed which scientist was pictured with which statement type), so that children did not respond on the basis of which scientist looked more knowledgeable. Materials were presented in one of the four different randomized orders that varied the sequence in which each item occurred.

Participants also completed an item at the end of each session that asked about distributed microstructure more generally. The experimenter said:

> Today in school, Molly and Erica learned about what makes an animal be a certain kind of animal. Their teacher told them about what makes a dog be a dog, a cat be a cat, and a mouse be a mouse. After school, Molly and Erica are trying to remember what they learned. Molly says that what makes an animal be a certain kind of animal is one special piece that is buried deep inside. Erica says that what makes an animal be a certain kind of animal is spread out through the entire body of animal.

Participants were asked to decide which person they thought was correct. The order of the distributed and localized statements, as well as which name was paired with which option, was balanced across participants.

Results

For each item, children's responses were classified as either distributed (i.e., *any location can identify the object*—coded as 1) or localist (i.e., *only one location can identify the object*—coded as 0). This produced five data points for each child (one for each item).

Two separate analyses were performed. We first examined the data nonparametrically. There were significant differences in the pattern of responses between the age groups for the "small animal," $\chi^2(2) = 7.52$, $p < .05$; "rock," $\chi^2(2) = 12.03$, $p < .01$; and "metal," $\chi^2(2) = 10.30$, $p < .01$ items. We then looked at each age group separately. Fourth graders were significantly more likely to choose the distributed option over the localized option for the "metal" and "rock" items ($p < .01$) and showed the same marginally significant pattern for the "large animal" and "kind-identity" items ($p = .06$) via binomial tests (two-tailed). Second graders' responses did not differ from chance (all $p > .38$). Kindergarteners, on the other hand, were more likely to choose the localized option over the distributed option for the "small animal" item ($p < .05$) via a binomial test (two-tailed), but did not show significant differences on any of the other items (all $p > .82$). These results are reported in Table 2.

In a second analysis, we summed the scores from all items to produce a score that ranged from 0 *localist response on all items*) to 5 (*distributed response on all items*); see the Appendix for a distribution of these parametric scores. A one-way ANOVA of this combined score comparing the three age groups revealed a significant main effect of grade on children's responses, $F(2, 64) = 9.86$, $p < .001$. Bonferroni post hoc analyses revealed that fourth graders were significantly more likely to endorse a distributed account ($M = 3.75$, $SE = .23$) than either second graders ($M = 2.43$, $SE = .27$), $p = .001$, or kindergarteners ($M = 2.20$, $SE = .32$), $p < .001$. Second graders and kindergarteners were not significantly different from one another. Comparisons to chance responding ($M = 2.5$), revealed that only fourth graders were significantly different than chance, $t(23) = 5.50$, $p < .001$.

Overall, children's justifications were too infrequent and sporadic to be coded and analyzed. However, only one child (age 10 years 7 months) used DNA to explain their response and a second child (age 10 years 4 months) mentioned cells.

Discussion

Results from Experiment 2 indicate that by at least fourth grade, children, like adults, tend to hold a distributed view of essence. These children reported that the unknown object's category could be determined by looking "anywhere," rather than in one specific location. Moreover, such beliefs do not appear to directly stem from knowledge about DNA or molecules because children hardly ever referred to these concepts. At the same time, fourth graders' responding was not identical across items. Endorsement of the distributed option for substances (rocks and minerals) was robustly above chance responding, whereas endorsement of the distributed option for animal items (small animal and large animal) only trended in that direction or was weakly significant. This pattern is interesting because it dovetails with the patterns observed in adults. Both adults and fourth graders tended to show weaker endorsement of the distributed option for animals compared to substances. As hypothesized before, this could be because a distributed understanding of complex, heterogeneous biological systems, such as animals, stems from a distributed understanding of less complicated, homogeneous structures, such as minerals. We return this issue at greater length in the General Discussion section.

The differences we found between fourth graders and younger children could be interpreted in one of the two ways: It could be that younger children simply do not have reliable intuitions about the physical nature of "very tiny internal pieces"; thus, they were at chance responding throughout. However, it may also be that younger children hold a consistent but competing view that we did not adequately measure with this task—namely, that *there is* one special piece that informs categorical identity better than others. Such an interpretation is at least plausible given that kindergarteners did favor this response at above chance levels for at least one item. Additionally, some of the younger children's justifications for selecting the localized option supported this view as well. For example, a few children spontaneously reported that the scientist would know best by looking "in the middle," whereas others offered structures such as "the blood" or "the heart." It could be that younger children really do tend toward a localist view of identifying internal properties, but the difficulty of the task obscured this pattern. In Experiment 3, we explored this issue further by creating a different scenario that simplified some of the task demands from Experiment 2.

Table 2
Proportion of Children in Each Age Group Choosing the Distributed Option in Experiment 2

Item	Kindergarten	Second grade	Fourth grade
Large animal	.55	.57	.71†
Small animal	.22*	.45	.63
Rock	.55	.39	.88**
Metal	.45	.48	.87**
Kind Identity	.50	.62	.71†

Note. Binomial test results (two-tailed).
$^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$.

Experiment 3

Results from Experiment 2 clearly suggested a developmental shift in children's endorsement of distributed views of essence. However, on the basis of that study alone, it was unclear exactly why younger children were at chance levels of responding. To address this question in Experiment 3, we employed a different but related measure to probe children's notions about essence. In this study, children were presented with two different kinds of objects that appeared the same on the outside. For example, children saw a picture of two lions that looked identical. However, they were told that one of the animals was a real lion, whereas the other was actually a tiger in a very good lion costume. Children were told that in order to figure out which was the real lion and which was the tiger, scientists looked at tiny pieces taken from the insides of the animals (see studies by Lizotte & Gelman, reported in Gelman, 2003, for a similar task in which children identify "insides" as vital to identifying the proper category). Unlike Experiment 2, however, we specified where the pieces were taken from—for example, from the middle of the animal and from the leg. This task was designed such that a distributed response would be to say that both pieces would tell you the same, whereas a localist response would be to say that a piece from one location might tell you better than a piece from another location. This task was in many ways superior to the task used in Experiment 2 because it eliminated any ambiguity about where the tiny pieces were taken from or how many their were. Moreover, this task better focused children's attention to "unseen," internal differences between the kinds, because the outward appearances of the objects were identical.

In this experiment, we also asked about artifacts to test the degree to which notions of distributed microstructure might be specific to animals and substances. Children saw two artifacts that looked identical on the outside, yet were different on one internal dimension (e.g., a pen and a mechanical pencil). For artifacts, a single functional component may indeed be the critical feature needed for categorization. Most artifacts are constructed of similar materials (e.g., plastics, various metals or alloys, wood, etc.), and rarely is it possible to distinguish an artifact solely on the basis of its *distributed* internal properties. Rather, these identifying markers of kind tend to be single functional units that in some way define or directly support that artifact's unique function. Thus, we predicted that children might have different intuitions about artifacts than animals and substances. For these items, all age groups should report that there is only one part (such as lead or ink) that can identify the object's category.

Method

Participants. A new group of 60 children participated in this study, including 20 kindergarteners (*M* age = 6.5), 20 second graders (*M* age = 8.5), and 20 fourth graders (*M* age = 10.5). Altogether there were 32 boys and 28 girls, roughly evenly distributed across grade. As in the previous study, participants were from middle- and upper-middle-class backgrounds and were recruited at public and private elementary schools. Ethnicity breakdowns were as follows: Caucasian ($n = 56$), Hispanic ($n = 1$), and African American ($n = 3$).

Stimulus materials and procedure. Participants were first shown a picture of two objects that appeared the same on the outside (see Figure 2). The experimenter said:

> Scientists find two animals. They know that one animal is a real lion and they know that the other animal is actually a tiger in a very good lion costume. Someone has played a trick and they have dressed up a tiger in a special skin-tight costume that makes it look just like a lion. The scientists cannot tell which is the real lion just by looking at them because they both look exactly the same on the outside. In order to figure it out, the scientists take really tiny pieces from the insides of the animals. Each piece is really, really, tiny. It is about the size of a pea. They will use all their tools and technology to look at the tiny pieces and they can see everything that is inside each piece. The scientists took a tiny piece from here in each of the animals (the middle), and the scientists took a tiny piece from here in each of the animals (the leg).

Participants were then asked:

> Which piece do you think will tell them the most about which is the real lion and which is the tiger? The piece from the middle, the piece from the leg, or will they both tell them the same?

For half of the participants, the middle piece was discussed first and for the other half of the participants, the leg piece was discussed first. This procedure was repeated for five additional pairs of items that consisted of one additional animal item (zebra/horse), two substance items (gold/coal and cement/wood), and two artifacts (pen/pencil and battery-operated watch/wind-up watch). The protocol was identical for each set of items, except for as follows: For substance items, the tiny pieces were taken either from the middle of the object or from a location that was off center. For artifact items, the tiny pieces were taken either from the middle of the object or from a nonfunctional location (the back of the pen or the band of the watch). Materials were presented in one of the four different randomized orders that varied the sequence in which each item was presented as well as which location (i.e., middle or other) was presented first.

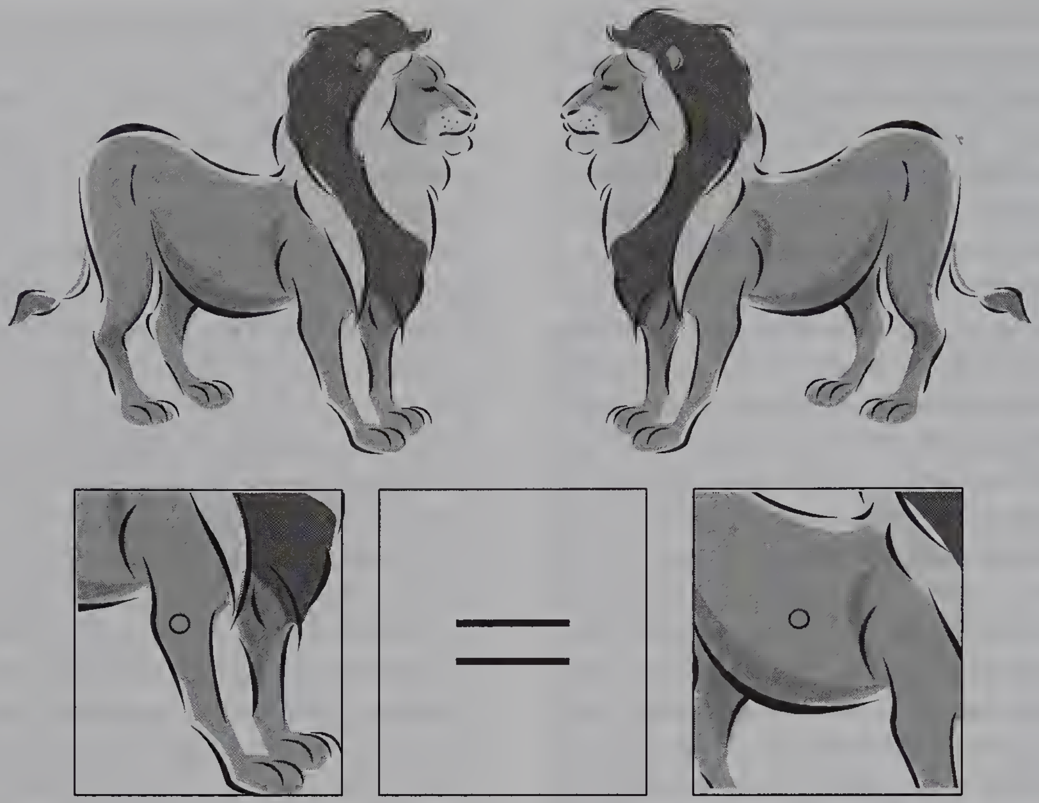


Figure 2. Sample pictures used in Experiment 3.

Results

Children's responses were classified as either *center* (e.g., "only the piece from the middle can identify the object") *other* (e.g., "only the piece from the inside of the leg can identify the object") or *both* (e.g., "both locations can identify the object"). The *center* and *other* options reflected a localist view of essence, whereas the *both* option reflected a distributed view.

The results from Experiment 3 are reported in Table 3. Consistent with the previous study, we found significant differences in the pattern of responses between age groups. We compared each of the grade levels using a Fisher's exact test and found significant differences between grades for the "gold/coal" ($p < .05$) and "cement/wood" ($p < .01$) items, and the same trend was present for the "lion/tiger" item ($p = .12$). There were no differences between grades for either of the artifact items (both $p > .24$). We then looked at each age group separately. Fourth graders were significantly more likely to say that pieces from *both* internal locations could identify "gold/coal," and "cement/wood" items ($p < .05$), and the "lion/tiger" item ($p = .055$) via binomial tests (two-tailed). As predicted, however, fourth graders were significantly more likely to say that pieces from "the center" could identify the "pen/pencil" and "watch" items ($p < .01$) via binomial tests. Additionally, because we predicted differences between natural kinds and artifacts, we

Table 3
Proportion of Children in Each Age Group Choosing Center, Other, *and* Both *Options in Experiment 3*

	Kindergarten			Second grade			Fourth grade		
Item	Center	Other	Both	Center	Other	Both	Center	Other	Both
Lion	.55*	.15	.30	.25	.30	.45	.40	.05	.55†
Zebra	.50	.28	.22	.30	.25	.45	.35	.15	.50
Cement	.60*	.20	.20	.40	.10	.50	.25	0	.75**
Gold	.75**	.10	.15	.32	.21	.47	.35	.10	.55*
Pen	.45	.50	.05	.70**	.30	0	.70**	.30	0
Watch	.90**	.10	0	.80**	.10	.10	.95**	0	.05

Note. Exact binomial test results (two-tailed).
$^{\dagger}p < .1$. $^{*}p < .05$. $^{**}p < .01$.

compared fourth graders' pattern of responding to each of the natural kinds items against the pattern of response to the two artifact items. Each of these comparisons was statistically significant via Fisher's exact tests (all $ps < .001$).

Second graders' pattern of response did not differ from chance for the natural kind items (all $p > .18$). However, similar to fourth graders, the second graders were significantly more likely to say that pieces from *the center* could identify the "pen/pencil" and "watch" items ($ps < .01$) via binomial tests. We then compared their pattern of response to each of the natural kind items against the artifact items via Fisher's exact tests. Each of these comparisons was statistically significant (all $ps < .001$).

In contrast, kindergarteners showed a very different pattern of response, as they tended to report that only the piece from *the center* could identify the natural kind objects. They were more likely to agree with the center option for the "gold/coal," and "cement/wood" items ($p < .01$), the "lion/tiger" item ($p = .055$), as well as the "watch" item ($p < .01$) via binomial tests (two-tailed). Unlike older children, kindergarteners' pattern of response to each of the natural kind items was not statistically different from the two artifact items, as they tended to endorse the "center" location option throughout.

Discussion

Experiment 3 found a developmental shift in how children think about the physical nature of essential features. Consistent with the previous experiment, fourth graders endorsed a distributed view of essence for animals and substances. In contrast, kindergarteners reported that for both animals and substances, the essence was localized to one specific region. And, second graders appeared to reflect a transition between these two understandings, as their responses were split between distributed and localized responses.

These data also helped resolve unanswered questions from the previous study. A concern from Experiment 2 was whether younger children really do show a preference for the localist view. Indeed, kindergarteners were significantly more likely to report that a "center" location would identify the objects for several of our items, even when given the option of another single, internal location. This pattern preliminarily suggests that kindergarteners actually appear to hold a rather specific belief about the physical nature of essence—namely, that it is located in the center of the object, for both animals and substances.

Data from the artifact items suggest that by at least fourth grade and possibly earlier, children make a sharp distinction between natural kinds and artifacts in terms of how they think essences are physically realized. For natural kinds, older children appreciate that essential features are distributed throughout the entity. However, for artifacts they recognize that there may be only one single location that correctly identifies the object. Kindergarteners, however, did not show this distinction, as they tended to favor a localist view throughout.

Finally, across all items, we tended to find greater agreement with the distributed view for substances than for animals. This result again parallels the results found in both Experiments 1 and 2. The appearance of such a pattern across three experiments is consistent with the interpretation that distributed beliefs about homogenous substances tend to emerge earlier than do distributed beliefs about animals and that such intuitions may, in fact, be used to buttress a distributed understanding of animals.

General Discussion

Children, as young as 6 years old, do not seem to be agnostic about the physical nature of essence. However, these younger children, contrary to adults, favor the view that essences are localized to the center of objects—not only for animals but for minerals as well. Around second grade, children begin shift away from this localist view to recognize that for minerals at least, essential features are distributed throughout. By fourth grade, children, like adults, recognize that for both minerals and animals, essential features are distributed—a view which they apply to natural kinds, but not to artifacts.

This shift from localist to distributed accounts is a developmental pattern that is distinct from what might be predicted by a "barest placeholder" notion of essence. We argue that instead, it may be incoherent to have a notion of essential features without some sense of how those features are physically instantiated. Thus, children and adults tend to constrain their inferences about essential features by making some rudimentary assumptions about where those features may be located: Initially, children seem to think of the essence as a single, localized thing, buried somewhere within the center. However, over time, this view is replaced with the notion that essential features are located throughout.

Not only do these results suggest that children may go beyond the barest placeholder notion of essence, but they also potentially speak to alternative accounts in the literature, which have argued that "blind-faith" essentialism based on barest placeholder notions alone may be implausible. For example, it has been suggested that children represent a few causal laws that link innards and heredity to phenomenal properties and that these laws masquerade as essentialist-like biases (Strevens, 2000). Along with others, we do not agree that such alternatives negate essentialism (see Ahn et al., 2001). Rather, we suggest that the results of the present studies point toward the idea that there is a drive to contextualize essentialism in terms of a more elaborate framework of causal assumptions—a framework that, for example, makes assumptions about how essences are physically instantiated. The critical modification here, however, is that this set of causal assumptions is far different from a detailed mechanistic understanding. It instead seems to consist of a very general sense of how and where causal features are located.

Indeed, the shift to a distributed view seems to occur without the benefit of detailed biological models or adult-like understandings of biology. Only one child in both studies ever mentioned DNA, and traditionally the concept of DNA is not introduced until sixth or seventh grade, at least 2 years older than children in our studies. Thus, even though a distributed view and knowledge about DNA make similar predictions about the location of identifying features (i.e., they can be found anywhere inside an entity), convictions about distributed essence seem to occur at a level of understanding far above that of concrete mechanism. The convictions instead seem to consist of a general sense of a causal pattern and its association with natural kinds. The presence of distributed reasoning in the absence of concrete theories about actual physical structures is additional support for the idea that children go beyond the barest placeholder notion of essence.

One potential limitation to these findings, however, concerns different sizes of microstructure. It is the case that the level of analysis suggested by our stimuli (e.g., pea-sized pieces) is larger than the level at which DNA is normally envisioned. And, for some adults, this may indeed result in different forms of reasoning. For children, however, we doubt that this difference was a factor. Moreover, attempts to realistically describe the appropriate microscopic scale would have not been feasible and well beyond the scope of this set of experiments. Thus, although we do not feel that this limitation prevents interpretation of these studies, it is an interesting question for future work as to how reasoning might vary as a function of the scale of the internal sample.

How might the shift from a localist to a distributed view of essence occur? One possible way may be through analogies to minerals, which seem to be understood as distributed somewhat earlier than living kinds. Experiment 1 found that adults were more likely to endorse the distributed account for minerals than they were for animals. Similarly, Experiments 2 and 3 found that the older children in our samples tended to endorse a distributed account for minerals more often than for animals. This pattern suggests a tendency to think that minerals are more likely to have distributed essential features than animals. We speculate that this contrast might be due to a number of factors. For example, real-world experience with substances such as rocks would reveal that their internal composition is homogenous. It may be that this type of real-world observation directly aided the distributed view in our tasks—for instance, leading children to respond that the scientists could take a piece from "anywhere" on the inside and still be able to identify the object's category. In turn, a distributed view of animals may occur through analogy to substances. Children may link knowledge about homogeneous distributed structures in substances to unseen distributed structures in animals. At the same time, though the data support the emergence of distributed reasoning for substances before animals, the hypothesis that real-world knowledge about substances actually "launches" later notions about living kinds needs to be confirmed in more focused experimental studies.

Moreover, this explanation does not address why younger children may hold a localized view of essence in the first place. A second, though not mutually exclusive, mechanism may be that the shift from localist to distributed accounts represents a more fundamental change in assumptions about the causal connection between internal causes and category membership. Previous work suggests that adults are often biased to prefer properties as central to categories when they are the single initial element in a causal chain (e.g., Ahn, 1999). It may be that younger children confuse a theoretical preference for a single, initiating cause with the physical existence of centrally located structures. This would be consistent with the finding that although young children insist that there is an identifying feature located in the center of objects, they never specify a particular internal part. It is also consistent with work showing that *what* children actually think the essence of an animal is, may change as a function of that child's particular culture (Waxman, Medin, & Ross, 2007). Moreover, in the case of substances, where young children express the same beliefs about a localized identifying feature, it is even less clear what actual physical structures could guide their intuitions, as there are none. It may be that the idea of a single initiating causal element is so compelling that it overrides considerable experiences with homogeneous substances suggesting otherwise.

We now return to the differences in adults' endorsement of the distributed statements between animals and substances. At the outset, we would have no reason to expect that adults should hold different views about animals and substances. However, we suggested that one explanation for adults' willingness to endorse Distributed views more often for minerals than for animals might be because more developmentally primitive ways of viewing these structures (i.e., localist notions) compete with the explicit knowledge that such features are distributed. To illustrate this point, consider the difficulty of reconciling Fact 1—that there is a single causal element, DNA, that causes biological kinds to be different—with Fact 2—that this same element is equally distributed around the body (represented in all nucleated cells). Although we speculate that most Western adults who have taken at least high school-level biology can confidently produce both of these facts, when it comes to explaining how DNA is then responsible for complex, synchronized developmental changes like growth or puberty, we speculate that most people's explanations will tend to fall flat—or more likely, revert to some sort of localist notion (e.g., that there is one special factor that oversees the process). Of course, this speculation requires empirical investigation, but examples such as this one raise important questions about what happens when even adult folk-biological models have to confront the true complexity of the biological world. We argue that people's lay models may have to be at least "a little localist" at times, simply because often the phenomena in question are so complex that they must be reduced to single-cause frameworks. Similarly, younger children, for reasons that are either specific to their particular theories of biology or constraints on their processing, may also be, in a sense, "required" to take a localist view of essential properties, merely because it represents a simpler causal story.

An additional mechanism that may promote these different views of essence could be some sort of instruction that is particular to a given culture. Would we see the same developmental patterns 150 years ago, before Mendel published his work on genetics—or even 200 years ago, before Brown observed the motion of microscopic particles? Or, would we expect the same patterns cross-culturally, in remote groups that might not be familiar with concepts such as DNA or atoms? Here, it is difficult to speculate, as the answers to these questions largely depend on how one believes children acquire notions of a distributed essence in the first place. On the one hand, if children develop such beliefs either through analogies to substances or a more profound change in their assumptions about underling cause, then we would expect the same types of patterns to occur, regardless of time or place. There is some support for this position, as ideas about atomistic or inherent, unseen causes far predate modern science and are culturally pervasive (Sober, 1994).

On the other hand, it may be that notions of a distributed microstructure reflect, or are in some way aided by our modern, Western scientific view. For example, work by Waxman et al. (2007) finds culturally specific intuitions about the nature of biological causes in children. They presented children from several different cultures with versions of the adoption paradigm in which children were told about a baby pig that was adopted and raised solely by other cows. Children were then asked about many of the pig's properties such as what sorts of things would be "pig like" and what sorts of things would be "cow like." Across all cultures, they found evidence for a strong biological component to children's understanding of property inheritance. Native American children, however, seem to have more firm commitments about the biological mechanism underlying the transmission of properties from parent to offspring. They often reported that the blood was essential, a belief that is widely shared among adults in this culture as well. The authors concluded that "children may be more likely to consider biological than nonbiological processes as candidate essences, and that in identifying candidate biological processes, they are sensitive to community-wide discourse" (p. 306). Thus, it is certainly plausible that children's notions of distributed essence are in some way shaped or buttressed by our modern scientific understanding and language.

In sum, we propose that even quite young children do not simply believe in a completely unspecified, barest placeholder notion of essence. In other words, they are not "blind-faith essentialists." Instead, they seem to embed essentialism into a more complex causal framework that takes into account the way in which essences may be physically instantiated. For natural kinds, children have some notions about the physical nature of essential features and will constrain their decisions about identifying features accordingly. At the same time, children may well be misguided early on, favoring a localized view of identifying features because it fits with a seemingly simpler causal story.

References

Ahn, W. (1999). Effect of causal structure on category construction. *Memory and Cognition*, *27*, 1008–1023.

Ahn, W., Kalish, C., Gelman, S. A., Medin, D. L., Luhmann, C., Atran, S., et al. (2001). Why essences are essential in the psychology of concepts. *Cognition, 82*, 59–69.

Appleman, P. (Ed.). (1979). *Darwin*. New York: Norton.

Au, T. K. F., Romo, L. F., & DeWitt, J. E. (1999). Considering children's folkbiology in health education. In M. Siegel & C. Peterson (Eds.), *Children's understanding of biology and health* (pp. 209–234). Cambridge, UK: Cambridge University Press.

Bloom, P. (2004). *Descartes' baby: How the science of child development explains what makes us human*. New York: Basic Books.

Descartes, R. (1972). *Treatise of man* (T. S. Hall, Trans.). Cambridge, MA: Harvard University Press. (Original work published 1664).

Dickinson, D. (1987). The development of a concept of material kind. *Science Education, 71*, 615–628.

Gelman, S. A. (1988). The development of induction within natural kind and artifact categories. *Cognitive Psychology, 20*, 65–95.

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. Oxford, UK: Oxford University Press.

Gelman, S. A., & Coley, J. D. (1990). The importance of knowing a dodo is a bird: Categories and inferences in 2-year-old children. *Developmental Psychology, 26*, 796–804.

Gelman, S. A., & Markman, E. M. (1987). Young children's inductions from natural kinds: The role of categories and appearances. *Child Development, 58*, 1532–1541.

Gelman, S. A., & Wellman, H. M. (1991). Insides and essences: Early understandings of the non-obvious. *Cognition, 38*, 213–244.

Gottfried, G. M., & Gelman, S. A. (2005). Developing domain-specific causal-explanatory frameworks: The role of insides and immanence. *Cognitive Development, 20*, 137–158.

Hull, D. (1965). The effect of essentialism on taxonomy: Two thousand years of stasis. *British Journal for the Philosophy of Science, 15*, 314–326, 16: 1–18.

Inagaki, K., & Hatano, G. (2002). *Young children's naive thinking about the biological world*. New York: Psychology Press.

Kalish, C. W., & Gelman, S. A. (1992). On wooden pillows: Multiple classification and children's category-based inductions. *Child Development, 63*, 1536–1557.

Keil, F. C. (1989). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.

Keil, F. C., & Richardson, D. C. (1999). Species, stuff, and patterns of causation. In R. Wilson (Ed.), *Species: New interdisciplinary essays*. Cambridge, MA: MIT Press.

Malt, B. C. (1994). Water is not H20. *Cognitive Psychology, 27*, 41–70.

Mayr, E. (1982). *The growth of biological thought: Diversity, evolution, and inheritance*. Cambridge, MA: Harvard University Press.

Medin, D. L., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical learning* (pp. 179–195). New York: Cambridge University Press.

Putnam, H. (Ed.). (1975). The meaning of 'meaning.' In *Mind, language, and reality: Philosophical papers* (Vol. 2). Cambridge, UK: Cambridge University Press.

Simons, D. J., & Keil, F. C. (1995). An abstract to concrete shift in the development of biological thought: The insides story. *Cognition, 6*, 129–163.

Sloman, S., & Malt, B. C. (2003). Artifacts are not ascribed essences, nor are they treated as belonging to kinds. *Language and Cognitive Processes, 18*, 563–582.

Sobel, D. M., Yoachim, C. M., Gopnik, A., Meltzoff, A. N., & Blumenthal, E. J. (2007). The blicket within: Preschoolers' inferences about insides and causes. *Journal of Cognition and Development, 8*, 159–182.

Sober, E. (1994). Evolution, population thinking, and essentialism. In E. Sober (Ed.), *Conceptual issues in evolutionary biology* (2nd ed., pp. 161–190). Cambridge, MA: MIT Press.

Springer, K., & Keil, F. C. (1991). Early differentiation of causal mechanisms appropriate to biological and nonbiological kinds. *Child Development, 62*, 767–781.

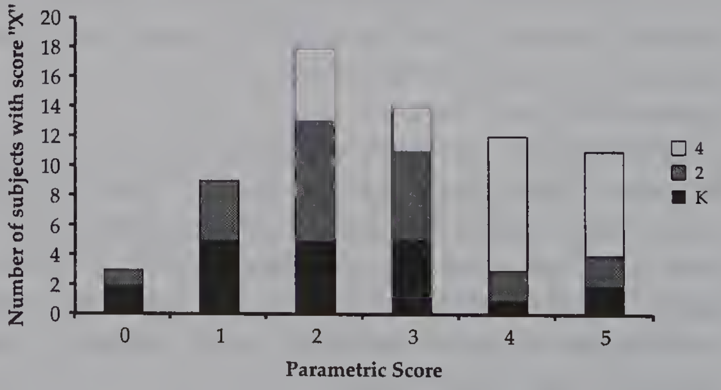
Stavy, R. (1991). Children's ideas about matter. *School Science and curriculum, 91*, 240–244.

Strevens, M. (2000). The essentialist aspect of naive theories. *Cognition, 74*, 149–175.

Waxman, S. R., Medin, D. L., & Ross, N. (2007). Folkbiological reasoning from a cross-cultural developmental perspective: Early essentialist notions are shaped by cultural beliefs. *Developmental Psychology, 43*, 294–308.

Appendix

Distribution of Parametric Scores in Experiment 2



Child Development, September/October 2008, Volume 79, Number 5, Pages 1357 – 1376

Profiles of Disruptive Behavior Across Early Childhood: Contributions of Frustration Reactivity, Physiological Regulation, and Maternal Behavior

Kathryn A. Degnan
The University of Maryland

Susan D. Calkins and Susan P. Keane
The University of North Carolina – Greensboro

Ashley L. Hill-Soderlund
The University of North Carolina – Chapel Hill

Disruptive behavior, including aggression, defiance, and temper tantrums, typically peaks in early toddlerhood and decreases by school entry; however, some children do not show this normative decline. The current study examined disruptive behavior in 318 boys and girls at 2, 4, and 5 years of age and frustration reactivity, physiological regulation, and maternal behavior in the laboratory at 2 years of age. A latent profile analysis resulted in 4 longitudinal profiles of disruptive behavior, which were differentiated by interactions between reactivity, regulation, and maternal behavior. A high profile was associated with high reactivity combined with high maternal control or low regulation combined with low maternal control. Results are discussed from a developmental psychopathology perspective.

Disruptive behaviors, such as aggression, defiance, and temper tantrums, are some of the most common problems seen in children (Beauchaine, Strassberg, Kees, & Drabick, 2002). Furthermore, chronic disruptive behavior problems are resistant to treatment and result in significant costs to society over time (Shaw et al., 1998). Given the negative effects and stability of disruptive behavior, it is important to identify its antecedents in order to develop useful early intervention and prevention programs. Typically, children's early social development includes a moderate level of disruptive behavior, which normatively desists (declines) across early childhood as early exposure to conflict helps children attain prosocial strategies (Owens & Shaw, 2003; Tremblay, 2000). However, some children retain their disruptive behavior across childhood, which often leads to antisocial behavior in adolescence and adulthood (Broidy et al., 2003). Understanding the precursors of persistent problem behavior may inform prevention efforts by elucidating the skills important to developing prosocial behavior and the risk factors that thwart these positive trajectories.

Consistent with a developmental psychopathology framework, theories of externalizing behavior suggest individual differences in developmental patterns of disruptive behavior. For example, past work has suggested two patterns of externalizing behavior, a high persistent pattern and a normative pattern (Moffitt, 1993). Additional research has demonstrated more than two longitudinal patterns of aggressive and nonaggressive externalizing behavior problems (Maughan, Pickles, Rowe, Costello, & Angold, 2000). Although typically defined a priori, these patterns can be empirically estimated using new statistical techniques, such as semiparametric mixture modeling (Muthén, 2004; D. S. Nagin & Tremblay, 2005). Studies using these techniques reveal empirical support for several patterns of disruptive behavior across early childhood (Hill, Degnan, Calkins, & Keane, 2006; National Institute of Child Health and Human Development [NICHD] Early Child Care Research Network, 2004; Shaw, Gilliom, Ingoldsby, & Nagin, 2003; Shaw, Lacourse, & Nagin, 2005). In addition, examining multiple longitudinal patterns of behavior,

---

This study was a dissertation project completed by the first author at The University of North Carolina at Greensboro. Support for this research was provided by a National Institute of Mental Health (NIMH) B/START award (MH 55625) and an NIMH FIRST award (MH 55584) to S.D.C., an NIMH grant (MH 58144) to S.D.C. and S.P.K., and an National Institute of Child Health and Human Development (HD 007376) predoctoral fellowship from the Center for Developmental Science to the first author. The authors would like to thank Daniel J. Bauer for his statistical consultation and Louise Berdan, Lauren Davis, Jessica Dollar, Keke Grady, Susan Owens, Cynthia Smith, and Michelle Felt for their help with data collection and coding. The authors also thank the families who generously gave their time to participate in the study.

Correspondence concerning this article should be addressed to Susan D. Calkins, Department of Human Development and Family Studies, Room 164C Stone Building, University of North Carolina at Greensboro, P.O. Box 26170, Greensboro, NC 27402-6170. Electronic mail may be sent to sdcalkin@uncg.edu.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7905-0010

rather than a single pattern or behavior at a single time point, will better represent the complexity of developmental pathways compared with more traditional approaches (Sroufe & Rutter, 1984).

The existence of multiple patterns of disruptive behavior has been supported; however, the array of factors that might distinguish these early chronic trajectories from more normative trajectories is less clear. Child and maternal factors, such as temperamental reactivity, emotion regulation, and maternal control, have been posited to influence children's ability to effectively control their emotions and learn socially appropriate behavior (Calkins, 1994; Calkins & Degnan, 2006). Previous research examined a few of these factors as direct predictors (Hill et al., 2006; NICHD Early Child Care Research Network, 2004; Shaw et al., 2005), and one study examined interactions of these factors in relation to longitudinal patterns of disruptive behavior in early childhood (see Shaw et al., 2003). However, from a developmental psychopathological perspective, further exploration into the joint effects of temperament and maternal behavior on longitudinal patterns of externalizing behavior problems is essential to fully inform prevention efforts. The concepts of multifinality and equifinality imply that both child and maternal factors may relate to specific levels of disruptive behavior problems in multiple ways (Cicchetti & Rogosch, 1996). Thus, the present study examined frustration reactivity, physiological regulation, and maternal controlling behavior in toddlerhood as joint predictors of the probability of membership in longitudinal profiles of disruptive behavior across early childhood.

Predictors of Longitudinal Profiles of Disruptive Behavior

One constellation of factors suggested to contribute to the maintenance of disruptive behavior is temperament, which includes reactivity to the environment and regulation of that reactivity (Cole, Martin, & Dennis, 2004). The stable and enduring aspect of temperament is thought to be an individual's level of emotional and motor reactivity during an emotionally challenging context, which is influenced by regulatory systems over time (Rothbart & Bates, 2006; Rothbart, Derryberry, & Hershey, 2000). In fact, observed measures of emotional reactivity during emotion-eliciting situations are moderately stable (Calkins, 2002; Stifter & Fox, 2000) and related to important behavioral outcomes (for a detailed review, see Rothbart & Bates, 2006). Whereas reactivity to novelty has been linked to internalizing behaviors (Fox, Henderson, Rubin, Calkins, & Schmidt, 2001), frustration reactivity has been linked to externalizing behaviors in toddlerhood and early childhood (Calkins, 2002; Hubbard et al., 2002). For instance, children who are highly reactive to frustration may act aggressively when provoked or aggravated because of their tendency to perceive these provocations as hostile (Vitaro, Brendgen, & Tremblay, 2002). In general, negative difficult temperamental components are associated with externalizing behavior problems (Rothbart & Bates, 2006). However, research exploring the effects of frustration reactivity on longitudinal patterns of problem behavior across early childhood is limited (see Shaw et al., 2005, for an exception).

Children's frustration reactivity is thought to be stable, but it is also influenced by regulatory systems over time (Rothbart & Bates, 2006; Rothbart et al., 2000). Thus, in addition to frustration reactivity, emotion regulation is also suggested to influence disruptive behavior. Whereas frustration reactivity is the emotional excitation of an individual, emotion regulation is the biological or behavioral process that alters that reactivity (Rothbart, Ahadi, & Evans, 2000). Through both physiological and behavioral factors, children first develop strategies to regulate arousal, which later develop into a formal repertoire of skills used to actively regulate emotions and behavior in a variety of contexts (Calkins, 1994; Calkins & Degnan, 2006). These skills lead to positive social skills and a decline in disruptive behavior over time (Denham & Burton, 1996). Emotion regulation is typically measured as behavior or physiology during situations that have presumed regulatory demands and elicit specific regulatory behaviors (Cole et al., 2004). A physiological construct frequently linked to emotion regulation is vagal tone (VT). When there are external demands, the autonomic nervous system supports the arousal of the sympathetic system by withdrawing the parasympathetic (vagal) input to the heart and other organs in order to promote fight/flight behaviors (Porges, 1991; Porges, Doussard-Roosevelt, & Maiti, 1994). Thus, under stress or challenge, respiratory sinus arrhythmia (RSA, an estimate of VT) is expected to decrease from baseline (vagal withdrawal; Porges et al., 1994), and this decrease is related to lower frustration reactivity, the use of emotion regulation skills (Calkins, 1997; Calkins, Dedmon, Gill, Lomax, & Johnson, 2002), and fewer behavior problems (Calkins & Dedmon, 2000; Eisenberg, Fabes, & Murphy, 1996).

Children's levels of frustration reactivity and physiological regulation have been directly related to externalizing behavior problems at specific time periods throughout childhood, but the longitudinal implications of early reactivity–regulation profiles are unknown. In addition, temperamental reactivity and emotion regulation are hypothesized to interact in relation to children's social adjustment (Calkins & Degnan, 2006; Rothbart & Bates, 2006; Sanson, Hemphill, & Smart, 2004). Undoubtedly, when children are able to regulate their frustration reactivity, they are better able to interact in a positive social manner. However, if children have limited regulatory skills and are prone to frustration, they are likely to develop maladaptive social behavior. Studies examining these types of interactive effects found that a combination of higher frustration reactivity and lower regulation skills predicts greater defiance (Stifter, Spinrad, & Braungart-Rieker, 1999), lower social competence (Belsky, Friedman, & Hsieh, 2001; Calkins, Gill, Johnson, & Smith, 1999), and greater externalizing behavior problems (Diener & Kim, 2004; Eisenberg et al., 2000). Also, an examination of observed and physiological measures of reactivity and regulation found that observed distress at 2 years of age was negatively related to 4-year positive social adjustment, but only for those children with lower physiological regulation (Calkins & Degnan, 2005). This work suggests that children's frustration reactivity and emotion regulation interact to predict disruptive behavior problems; however, few studies have explored these interactions in relation to longitudinal patterns of behavior. From a developmental psychopathology viewpoint, examining measures of reactivity and regulation from multiple levels of analysis as joint influences on longitudinal pathways of behavior problems would support the value of these skills for the development of positive social behavior (Kuperminc & Brookmeyer, 2006). Therefore, the current study examined the joint effects of observed frustration reactivity and physiological regulation in toddlerhood in relation to longitudinal patterns of disruptive behavior in early childhood.

Another factor that influences children's levels of frustration reactivity, emotion regulation, and socially appropriate behavior is maternal behavior. From birth, caregivers assist infants with general state regulation by providing basic necessities (e.g., food and clothing); however, during the transition to toddlerhood, this assistance evolves into more complex social interactions in which children learn to manage their own distress and behavior. The development of these self-management abilities is expected to occur with the support of a positive mother – child relationship (Calkins, 1994). In contrast, children's aggression, noncompliance, or broader externalizing problems are associated with controlling, rejecting, or harsh parenting behavior (Campbell, Pierce, Moore, Marakovitz, & Newby, 1996; Rubin, Hastings, Chen, Stewart, & McNichol, 1998; Shaw et al., 1998).

Maternal behaviors, such as control or rejection, may also interact with children's frustration reactivity and emotion regulation in relation to disruptive behavior problems (Calkins & Degnan, 2006; Rothbart & Bates, 2006). For instance, maternal behavior may enhance or undermine the development of important regulatory skills, which reactive children would have difficulty acquiring otherwise (Braungart & Stifter, 1991; Calkins, 2002; Calkins, Smith, Gill, & Johnson, 1998). Specifically, mothers who are sensitive and responsive may help children who have high frustration reactivity and regulatory difficulties develop appropriate social behavior. In contrast, mothers who are intrusive or hostile may exacerbate their children's frustration reactivity and poor regulatory skills, leading to greater disruptive behavior problems (Gilliom, Shaw, Beck, Schonberg, & Lukon, 2002; Rubin, Burgess, Dwyer, & Hastings, 2003). Supporting this notion, research examining the joint effects of maternal behavior and temperament reveals that children's frustration reactivity typically predicts externalizing outcomes when mothers are high on control or low on positivity (Calkins, 2002; Paterson & Sanson, 1999; Rubin et al., 2003). However, one study found that maternal control had a positive impact on children's frustration reactivity (Bates, Pettit, Dodge, & Ridge, 1998), suggesting that the same level of maternal control may not operate the same way in all contexts or with all children. In some instances (e.g., when danger is present; Bates et al., 1998), maternal control may help dampen children's frustration reactivity and limit the display of disruptive behavior. However, in other situations (e.g., free play; Calkins, 2002), maternal control may exacerbate a negative temperament and increase disruptive behavior. In addition, maternal control may lead to disruptive behavior when children are not prone to frustration but have strong approach tendencies. Parents of this type of child may attempt to inhibit the approach behavior, frustrate the child, and increase the likelihood of disruptive behavior (Derryberry & Reed, 1994). Overall, maternal controlling behavior may have different implications for child externalizing outcomes depending on the child's individual level of frustration reactivity and regulatory capabilities. Therefore, the current study examined the joint effects of maternal control, child observed frustration reactivity, and child physiological regulation in toddlerhood in relation to longitudinal patterns of disruptive behavior in early childhood.

Goals and Hypotheses

Consistent with the goal of developmental psychopathology, this study aimed to examine multiple longitudinal patterns of disruptive behavior while investigating the joint effects of both child and maternal factors on these multiple patterns of development. The primary goal of this study was to use a semiparametric group-based approach to examine latent profiles of disruptive behavior from 2 to 5 years of age. Multiple longitudinal patterns of disruptive behavior across childhood were expected due to individual differences in the initial propensity for externalizing behavior problems and the development of socially appropriate behavior over time (Moffitt, 1993; Owens & Shaw, 2003). Similar to those found in previous research on early childhood (e.g., Hill et al., 2006; Shaw et al., 2003), four profiles of disruptive behavior were hypothesized to emerge: a high profile; a high, declining profile; a moderate, declining profile; and a low profile.

The second goal of this study was to examine the joint effects of observed child frustration reactivity, physiological emotion regulation, and observed maternal controlling behavior in toddlerhood on the probability of membership in the profiles. Toddlers prone to high frustration reactivity were hypothesized to exhibit high levels of disruptive behavior; however, physiological emotion regulation was posited to protect children from maintaining these behavior problems across early childhood. In addition, the level of maternal control was expected to be positively related to profiles with greater disruptive behavior problems in general, but especially when children displayed higher rates of frustration reactivity and lower physiological regulation. Although these factors (frustration reactivity, physiological regulation, and maternal control) have been directly related to disruptive behavior problems in previous work, the present study examined how they might interact to differentiate multiple profiles of disruptive behavior over early childhood. Although this approach was somewhat exploratory, specific interaction effects were expected to emerge. Specifically, high levels of frustration reactivity, lower levels of physiological regulation, and higher maternal control were hypothesized to increase the likelihood of membership in the high disruptive behavior profile. In contrast, the probability of membership in the high, declining profile was expected to be related to a combination of higher levels of both frustration reactivity and physiological regulation supported by lower levels of maternal control. The probability of membership in the moderate, declining and low profiles was expected to be related to lower levels of frustration reactivity, higher levels of physiological emotion regulation, and lower levels of maternal control. However, a higher probability of membership in the moderate, declining profile was expected to be related to slightly more frustration reactivity than the probability of membership in the low profile. In summary, child frustration reactivity, physiological emotion regulation, and maternal control were expected to collectively differentiate the probability of membership in the disruptive behavior profiles.

Method

Participants

Participants for the current study were part of a longitudinal study of four hundred and forty-seven 2-year-old children (215 boys and 232 girls) obtained from three cohorts. Sixty-seven percent were European American, 27% were African American, 4% were biracial, and 2% were Hispanic. At age 2, the children were primarily from intact families (77%) who were economically diverse with Hollingshead (1975) scores ranging from 14 (unskilled laborers/menial service workers) to 66 (major business professionals), with an average score of 39.56 (medium business professionals). Parental education levels ranged from some high school (3% mothers and 6% fathers) to advanced degrees (12% mothers and 11% fathers), and over 30% of mothers and fathers had completed college.

Recruitment

The goal for recruitment was to obtain a representative community sample of children, some of whom were at risk for developing future externalizing behavior problems. Thus, all cohorts were recruited through child day care centers, the County Health Department, and the local Women, Infants, and Children program. The sample was screened by gender to obtain equal numbers of males and females. In addition, recruitment was targeted to all areas of the county to obtain a sample representative of the surrounding area in terms of race and socioeconomic status (SES).

Potential participants for Cohorts 1 and 2 ($n = 307$) were recruited at 2 years of age (Cohort 1: 1994 – 1996 and Cohort 2: 2000 – 2001) and screened using mother report on the Child Behavior Checklist (CBCL 2 – 3; Achenbach, 1992). Children with an externalizing T score of 60 or above were selected to be in the *externalizing risk* group ($n = 143$). Those with both externalizing and internalizing T scores below 60 were selected to be in the *low-risk* group. Cohort 3 was initially recruited when infants were 6 months of age (in 1998) for their level of frustration based on laboratory observation and parent report (see Calkins et al., 2002, for more information). This cohort was followed from 6 months of age through the infancy and toddler period, and children whose mothers completed the CBCL at 2 years of age were included in the current study (*n* = 140). Based on the criteria described earlier, 21 children from this cohort were placed in the externalizing risk group. Cohort 3 had a significantly lower average 2-year externalizing *T* score (*M* = 50.36) compared to Cohorts 1 and 2 (*M* = 54.49), *t*(445) = −4.32, *p* = .00. Of the entire sample (*N* = 447), 164 children met criteria for the externalizing risk group. There were no significant differences between any cohorts with regard to gender, $\chi^2(2, N = 447) = 0.63, p = .73$; race, $\chi^2(2, N = 447) = 1.13, p = .57$; or 2-year SES, *F*(2, 444) = 0.53, *p* = .59.

Missing Data

Of the 447 participants originally selected at 2 years of age from maternal report on the CBCL, 6 were dropped because they did not participate in any 2-year laboratory data collection. At 4 years of age, 399 families participated. Families missing at 4 years of age included 20 who could not be located, 10 who moved out of the area, 9 who declined participation, and 9 who did not respond to phone and letter requests to participate. There were no significant differences between families who did and did not participate at 4 years of age in terms of gender, $\chi^2(1, N = 447) = 3.27, p = .07$; race, $\chi^2(1, N = 447) = 0.70, p = .40$; 2-year SES, *t*(424) = 0.81, *p* = .42; or 2-year externalizing *T* score, *t*(445) = −0.36, *p* = .72. At 5 years of age, 365 families participated including 4 who did not participate in the 4-year assessment. Families missing at 5 years of age included 12 who could not be located, 10 who moved out of the area, 13 who declined participation, and 3 who did not respond to phone and letter requests to participate. Again, there were no significant differences between families who did and did not participate at 5 years in terms of gender, $\chi^2(1, N = 447) = 0.76, p = .38$; race, $\chi^2(1, N = 447) = 0.17, p = .68$; and 2-year externalizing *T* score, *t*(445) = −1.73, *p* = .09. There was a trend for a difference between the families who did or did not participate at 5 years in terms of 2-year SES, *t*(424) = 1.93, *p* = .06, such that families with lower SES were not as likely to participate, although this effect size was minimal (*d* = 0.22).

Procedures

Two-Year Assessment

Mothers brought their children to the laboratory and were videotaped during several episodes designed to elicit frustration reactivity and mother–child interaction. All episodes were videotaped for later coding. If the child became highly distressed (i.e., cried hard for more than 30 s), the episodes were ended early. In addition, mothers were asked to complete multiple questionnaires including the CBCL 2–3 (Achenbach, 1992) and a demographic measure. At the end of the laboratory visit, mothers were compensated for their time and children were given a small prize for their participation.

RSA assessment. At the beginning of the 2-year laboratory visit, an experimenter placed three disposable, pediatric electrodes in an inverted triangle pattern on the child's chest. The electrodes were connected to a preamplifier, and the output from the preamplifier was transmitted to a vagal tone monitor (VTM-I; Delta Biometrics, Bethesda, MD) for R-wave detection. The vagal tone monitor displayed heart rate throughout the baseline, toy/cookie in box, and teaching episodes, and every 30 s, it computed and displayed RSA values. A data file containing the interbeat intervals (IBIs) for the entire period of heart rate collection was saved on a laptop computer for later artifact editing (e.g., resulting from child movement) and analysis. For the current study, only RSA measures during the baseline and first frustration (toy/cookie in box) episodes were used.

For the baseline episode, the child watched a 5-min segment of the videotape "Spot," a story about a puppy that explores its neighborhood. Although this episode was not a true baseline, as the child's attention was engaged, it was sufficient to gain a measure of RSA while the child was sitting quietly and showing neutral affect. Given these children were 2 years of age, such a stimulus was necessary in order to limit movement artifact in the heart rate data. Following the baseline episode was the first frustration episode (toy/cookie in box) during which the children were asked whether they wanted a snack or to play with an exciting toy. If the child was asked to play with a toy, they were permitted to play with the toy for 1 min. Then, the experimenter placed the snack or toy in a clear plastic container that the child could see and touch, but was unable to open, for 2 min. Throughout this episode, the mother was nearby and was instructed to respond to the child as she normally would but to limit initiating any interaction. This task was adapted from Laboratory Temperament Assessment Battery procedures (Lab-TAB; Goldsmith & Rothbart, 1996). Next, the mother was instructed to teach the child how to complete a challenging puzzle or shape sorter for 3 (Cohort 3) or 4 (Cohorts 1 and 2) min. When the RSA assessment period was complete, the electrodes were removed from the child's chest and the child was given a snack break.

Laboratory assessment. Following the break, the second frustration task (high chair) was administered. The child was placed in a high chair, without any toys or snacks, for 5 min. Throughout this episode, the mother was seated nearby and was instructed to respond to her child as she deemed necessary. At the end of the 5 min, the child was removed from the high chair. Next, a farm or pretend town was provided and the mother was instructed to play with her child as she would at home for 4 min. Following the free play episode, the mother was instructed to get the child to clean up the toys as they would at home. This episode lasted 2 min or until all the toys were put away, whichever happened first.

Four- and 5-Year Assessments

The mothers were requested to accompany their children to the laboratory when they were 4 and 5 years of age. During these visits, mothers completed multiple questionnaires including the CBCL 4–18 (Achenbach, 1991a). For each visit, mothers were compensated for their time and children were given a small prize for their participation. For the purposes of the current study, only the CBCL data from these visits were examined.

Measures

Demographics

Mother's open response report of children's race and gender was used to measure children's race (Caucasian or minority) and gender (male or female). Mothers also reported their own and the child's father's (if he was contributing to the household) education level, marital status, and type of employment. These measures were used to construct an SES score for the 2-year laboratory visit based on the Hollingshead (1975) index. This score is derived from maternal and paternal education level and occupational strata.

Disruptive Behavior Problems

The CBCL's (Achenbach 1991a, 1992) aggression subscale was used as an index of mother-reported disruptive behavior problems at each age. When the children were 2 years of age, mothers completed the CBCL for 2- to 3-year-olds (Achenbach, 1992). When the children were 4 and 5 years of age, mothers completed the CBCL for 4- to 18-year-olds (Achenbach, 1991a). These scales are a reliable index of various externalizing and internalizing behavior problems across childhood (Achenbach, Edelbrock, & Howell, 1987). Both versions included a narrow-band *aggression* subscale, which consisted of items measuring physical aggression, anger, and general cruelty toward others. At 2 years of age, the subscale consisted of 15 items such as "defiant," "fights," and "hits others." At 4 and 5 years of age, the subscale consisted of 20 items such as "argues," "mean to others," and "physically attacks people." The mother indicated how true each item was of her child by circling 0 if *not true*, 1 if *sometimes true*, or 2 if *often true*. Although Achenbach (1991a, 1992) labeled this factor aggression, an examination of the items suggests that it is measuring a somewhat broader construct of disruptive behavior, including defiance, destructiveness, and physical aggression (Beauchaine et al., 2002). Thus, the term *disruptive behavior* will be used with regard to this measure.

Although the CBCL includes *T* scores for each subscale, for the purposes of this study, the total scores of the subscales were used in order to allow for maximum variation across the sample with a possible range from 0 to 30 for the measure at 2 years of age and 0–40 for the measures at 4 and 5 years of age. It is noted that the range and items of each scale are different because they measure identifiable and expected disruptive behaviors for either the 2- to 3- (15 items) or 4- to 18-year old (20 items) age range. On average, for this sample, the 4- and 5-year-olds' scores were lower than the 2-year-olds' scores (Table 1). In order to control for the different number of items at each age, the mean of the subscales was created by dividing the raw scores by their respective number of items, creating a possible range of scores between 0 and 2 at each age. Averages and standard deviations of the disruptive behavior total raw and mean scores are in Table 1.

Observed Frustration Reactivity

Reactivity behaviors were coded from videotapes of both frustration episodes (toy/cookie in box and high chair). Reactivity was indexed by measures of distress, including whining, pouting, fussing, crying, screaming, or tantrumming. It was coded in three ways: (a) *proportion of distress*: the amount of time (in seconds) the child was distressed divided by the total time of the episode; (b) *global negative reactivity*: coded for the entire episode on a scale from 0 (*no negative response*) to 4 (*episode ended with the child in extreme distress*); and (c) *global episode affect*: coded once for the entire episode on a scale from −3 (*highly distressed affect*), to 3 (*highly positive affect*). These measures were thought to best index a child's level of observable

Table 1
Descriptive Statistics

Measures	*N*	Minimum	Maximum	*M*	*SD*
2-year socioeconomic status	318	14.00	66.00	39.56	10.82
2-year proportion of distress	318	0.00	0.98	0.11	0.17
2-year global negative affect	318	0.00	4.00	0.75	0.85
2-year global episode affect	318	−3.00	2.00	−0.11	0.78
2-year observed reactivity score	318	−1.26	1.90	0.00	0.43
2-year baseline RSA	318	1.10	8.79	5.49	1.27
2-year frustration RSA	318	1.45	8.57	4.89	1.16
2-year physiological regulation score	318	−3.37	3.34	0.00	1.00
2-year adult-centered statements	318	6.67	41.11	20.03	5.95
2-year global strictness/punitiveness	318	1.00	4.00	2.48	0.64
2-year maternal control score	318	−2.01	2.43	0.00	0.96
2-year disruptive behavior total	318	0.00	29.00	10.04	5.68
2-year mean disruptive behavior score	318	0.00	1.93	0.67	0.38
4-year disruptive behavior total	274	0.00	28.00	8.93	5.97
4-year mean disruptive behavior score	274	0.00	1.40	0.45	0.29
5-year disruptive behavior total	252	0.00	33.00	8.73	6.84
5-year mean disruptive behavior score	252	0.00	1.65	0.44	0.34

Note. RSA = respiratory sinus arrhythmia.

frustration reactivity during episodes with presumed regulatory demands (Calkins, 1997; Stifter & Braungart, 1995). Four coders trained on 10% of the videotaped sessions and independently coded another 10% for reliability. Intercoder reliability for the proportion of distress measure was excellent, mean $r = .99, p < .00$ (range = .98 – 1.0). Reliability kappas for the ordinal codes were .83 for global negative affect and 1.0 for global episode affect. Each of the reactivity codes was correlated and averaged across episodes (average $r = .25$, $p < .00$). Descriptive statistics for each average code are reported in Table 1.

In order to compute an "observed reactivity" composite, the average codes were standardized and summed ($\alpha = .86$). This composite had a high skewness value of 1.94 and a kurtosis value of 5.53. Therefore, the composite was transformed using a square root transformation bringing the skewness value to 1.19 and the kurtosis value to 2.25. Then, the transformed composite score was centered to allow for ease of interpretation in the model. The mean and standard deviation for this final composite variable are reported in Table 1.

Physiological Emotion Regulation

Measures of children's RSA during the baseline and toy/cookie in box episodes were obtained by editing the IBI files using MXEDIT software (Delta Biometrics, Bethesda, MD.). The Porges' (U.S. Patent No. 4520944, 1985) method of analyzing the IBI data was used to calculate RSA. This method applies an algorithm to the sequential heart period (HP) data. The algorithm uses a moving 21-point polynomial to detrend periodicities in HP that are slower than RSA. Then, a band-pass filter extracts the variance in HP within the frequency band of spontaneous respiration in young children, 0.24 – 1.04 Hz. The natural log of this variance is taken and reported in units of $\ln(\text{ms})^2$. To edit the files, the data were scanned for outlier points, relative to adjacent data, and the outliers were replaced by dividing or summing them so they would be more consistent with the surrounding data. Only data files in which less than 10% of the data required editing were included in the current study. For the baseline and toy/cookie in box episodes, RSA was calculated every 30 s and averaged across epochs. Individual data were excluded if the standard deviation for an episode was over 1.0 (see MXEDIT manual; U.S. Patent No. 4520944, 1985).

Measures of RSA during the baseline and toy/cookie in box episodes were used by calculating a difference score (Calkins, 1997; Moore & Calkins, 2004) to create an index of "physiological regulation" during the frustration episode, thus taking into account the level of baseline RSA. This was done by subtracting RSA during the frustration episode from RSA during the baseline episode. Positive change scores occurred when there was a decrease from the baseline to the frustration episode, which reflects attempts to regulate emotion. Negative change scores occurred when there was an increase from the baseline to the frustration episode, which reflects a lack of physiological regulation. The RSA change score was then centered to allow for ease of interpretation in the model. Descriptive statistics for the measures of RSA during baseline and frustration episodes and the physiological regulation score are reported in Table 1.

Maternal Behavior

Maternal behavior, during the interactive episodes (teaching, free play, and cleanup), was coded with two coding systems. One system (Rubin, Booth, Rose-Krasnor, & Mills, 1995; Smith, Calkins, Keane, Anastopoulos, & Shelton, 2004) recorded the implied goal of each maternal statement: *Adult-oriented* goals included initiating a new activity (a beginning) or stopping the child's activity (a stop), as opposed to continuing or facilitating the child's own activity. The duration of episodes varied across dyad, so the sum of adult-oriented statements (beginnings and stops) was standardized for each participant by dividing the total number by the total time of the individual's episode and multiplying this value by the maximum time of the episode (teaching episode: 4 min, free play episode: 4 min, and cleanup episode: 2 min). Four coders trained on 10% of the videotaped sessions and independently coded another 10% for reliability. The average intercoder reliability for the maternal goal measures was $r = .86$, $p < .00$ (range = .81 – .88).

The second coding system (Smith et al., 2004; Winslow, Shaw, Bruns, & Kiebler, 1995) examined a global index of *strictness/punitiveness* (being too strict, demanding, or harsh relative to the child's behavior; exerting influence toward completion of the child's activity; displaying a no-nonsense attitude; and constantly guiding the child and creating a very structured environment). This was coded once for each episode on a 4-point scale, ranging from low levels of the behavior to high levels of the behavior. Four coders trained on 10% of the videotaped sessions and independently coded another 10% for reliability. The reliability kappa between each pair of coders was .71 for strictness/punitiveness.

The adult-oriented goals and the global code for strictness/punitiveness were each averaged across episodes (average $r = .44$, $p < .00$; average $\alpha = .58$). Descriptive statistics for each average code are reported in Table 1. Following Smith et al. (2004), a composite "maternal control" was created by standardizing and summing the average adult-oriented goals and global strictness/punitiveness measures ($\alpha = .76$). This composite was then centered to allow for ease of interpretation in the model. The mean and standard deviation for this maternal control measure are reported in Table 1.

Summary of Measures

The observed reactivity, physiological emotion regulation, and maternal control measures at 2 years of age and the CBCL (Achenbach, 1991a, 1992) subscales of aggression at 2, 4, and 5 years of age were examined in the present study. Of the possible 441 participants, only 318 (72%) had complete data at 2 years of age: Ninety-seven had technical difficulties with the physiological data collection, 12 could not be contacted by phone to schedule a laboratory visit but completed questionnaires through the mail, 11 had technical difficulties with the video equipment, and 3 refused to complete socioeconomic information. These families were not significantly different from the overall sample by gender, $\chi^2(1, N = 447) = 0.38$, $p = .54$; race, $\chi^2(1, N = 447) = 0.04$, $p = .85$; 2-year SES, $t(424) = 0.00$, $p = 1.0$; or 2-year externalizing T score, $t(445) = -0.21$, $p = .84$. Of these 318 participants, 274 had complete 4-year CBCL data and 252 had complete 5-year CBCL data. These families were not significantly different from those with incomplete CBCL data by gender, $\chi^2(1, N = 318) = .1.49$, $p = .22$; race, $\chi^2(1, N = 318) = 0.99$, $p = .32$; or 2-year externalizing T score, $t(316) = -1.35$, $p = .18$. However, there was a trend for those with missing data to have significantly lower 2-year SES, $t(313) = 1.95$, $p = .05$. Further analysis determined that the patterns of missing data at 4 and 5 years of age were missing completely at random (MCAR), Little's MCAR test, $\chi^2(28, N = 318) = 16.26$, $p = .96$. In addition, the present analysis used maximum likelihood estimation, which allows for missing data longitudinally and assumes the missing data across repeated measures are missing at random (Little & Rubin, 1987; Muthén & Muthén, 2006). Thus, despite missing data across time, 318 families provided data for the current study, 124 (39%) with children above the borderline clinical range on 2-year externalizing behavior and 194 (61%) with children below the borderline clinical range on 2-year externalizing behavior.

Data Analyses Goals

To investigate individual differences in longitudinal patterns of disruptive behavior, a structural equation mixture model (SEMM) was used. As a semiparametric group-based approach, SEMM allows for estimation of qualitatively different groups (i.e., classes) when group membership cannot be observed a priori (Bauer & Curran, 2004). Recent work of D. Nagin and Tremblay (1999) and Muthén (2001) showed how SEMMs can be used in testing differential longitudinal patterns of psychological phenomena. In the current study, disruptive behavior at age 2 was measured with a different form of the CBCL (age 2–3) than disruptive behavior at ages 4 and 5 (CBCL 4–18). Thus, linear growth trajectories of disruptive behavior were not estimated (e.g., D. Nagin & Tremblay, 1999) due to the possibility of change in measurement; rather, the average level of disruptive behavior at each age was estimated independently within each class (i.e., latent profile analysis [LPA]; Gibson, 1959). For this reason, classes are referred to as "longitudinal profiles" rather than trajectories. In this study, the longitudinal profiles described levels of disruptive behavior at 2, 4, and 5 years of age.

As mentioned, one benefit of using an SEMM model such as LPA is that it performs maximum likelihood estimation, which includes all longitudinal observations in a data set (Little & Rubin, 1987). This method assumes the data are missing at random and has been recently recommended by methodologists as an appropriate way to accommodate missing data (Schafer & Graham, 2002). Using this method of estimation allows the model parameters to be informed by all cases that contribute a portion of the data. As a submodel of SEMM, LPA is a multiple group structural equation model in which the group variable is unobserved. Thus, LPA assumes observed associations are explained by differences in the means of the continuous measures over latent classes (Bauer & Curran, 2004). With this framework, individual's probabilities of membership in each profile and the predictor's effects on those probabilities are estimated in the same analysis. As with a multinomial logistic regression, the analysis compares the probability of membership in all the profiles in reference to one of the profiles (e.g., 1, 2, 3 vs. 4). In addition, the statistical package used (Mplus 4.1) estimated all possible comparisons in the same analysis (i.e., tests all the other profiles as the reference profile).

Although similar to logistic regression and cluster analysis, LPA differs from past methods used to identify groups in two ways. First, LPA relies on a formal statistical model rather than an ad hoc algorithm based on decision rules (e.g., cluster analysis) and allows for flexibility in the model (Everitt & Hand, 1981). Second, the flexibility of model-based LPA allows for the possibility of uncertainty in which classes people may belong to, allowing one to predict the *probability* of membership in a group (rather than *membership* in that group per se). In other words, unlike cluster analysis, people are not forced into a group so that additional analyses can be performed to examine predictors; rather, all analyses are performed within one formal statistical model.

The function for LPA takes the general form as follows:

$$Y_{(tik)} = \mu_{(tk)} + \varepsilon_{(tik)},$$

where $\mu_{(tk)}$ is the class-specific mean for the observed variable Y at time t for class k, and $\varepsilon_{(tik)}$ are within-class individual differences from $\mu_{(tk)}$. Within each class, $\varepsilon_{(tik)}$ is assumed to be normally distributed with variance $\sigma_{(tk)}$, allowing for potential heteroscedasticity across time and classes. In this study, the Y variables are disruptive behavior scores at ages 2, 4, and 5, and the estimated class means $\mu_{(tk)}$ for these variables describe the longitudinal latent profile for each class.

In the current study, data were analyzed using Version 4.1 of Mplus (Muthén & Muthén, 2006), and models with two through six profiles were estimated. Determination of best model fit was assessed using Bayesian information criteria (BIC), where the smallest negative number indicates best fit. This index has been shown to identify the appropriate number of groups in finite mixture models (D'Unger, Land, McCall, & Nagin, 1998) and penalizes the model for the number of parameters, thus guarding against models overfitting the data. In addition, when fitting models like these, issues such as convergence are important, especially for more complex models (Hipp & Bauer, 2006), and sometimes, start values must be specified in order to reach convergence. For this case, the model was relatively simple, and random start values resulted in a converged solution. Predictors of the probabilities of membership in the above profiles were entered into the model in a hierarchical fashion. The log likelihoods of each step were compared with the prior step to determine whether the predictors significantly affected the model (McArdle, 2005). If they did not affect the model, then they were taken out to conserve power needed to detect interaction effects. Additionally, interaction effects were probed similarly to how one would test interactions in logistic regression using the guidelines of Aiken and West (1991).

Results

Descriptive Analyses

A two-way multivariate analysis of variance, with race (Caucasian vs. minority) and gender, was used to test for group differences on all outcomes variables (disruptive behavior at 2, 4, and 5 years of age). There were no significant main effects of race or gender on the outcome variables. In addition, follow-up analyses indicated that the demographic factors and their moderation effects did not significantly impact the latent class analysis above and beyond the effects of the main predictors. Thus, in subsequent analyses, gender and race are not included in the model. Correlations between 2-year SES, baseline RSA, and all outcome variables (disruptive behavior at 2, 4, and 5 years of age) revealed a single significant association. Family SES at 2 years of age was negatively related to disruptive behavior at 2 years of age ($r = -.17, p < .01$). Thus, children rated by their mothers as displaying more disruptive behavior at age 2 also had parents with lower average SES at age 2; subsequently, SES was tested in the model. Baseline RSA was not significantly related to the measures of disruptive behavior and was not controlled for in the model. Finally, correlations were computed between the predictor variables themselves (frustration reactivity, physiological regulation, and maternal control). Maternal control was not significantly associated with frustration reactivity or physiological regulation; however, frustration reactivity and physiological regulation were modestly intercorrelated, $r = .15, p = .01$.

Latent Profile Model Comparisons

Latent profile models with two through six profiles were fit to determine the optimal number of profiles to describe disruptive behavior from 2 to 5 years of age in the current sample. Model fit was assessed using the BIC, where the smallest negative number (closest to 0) indicates best fit. The BIC for the current sample was −311 for two profiles, −248 for three profiles, −238 for four profiles, −244 for five profiles, and −275 for six profiles. The four-profile model had the smallest negative BIC and therefore was selected as the best number of disruptive behavior profiles for the current sample. As a secondary test of model fit, the Lo – Mendell – Rubin likelihood ratio test was also used, which tests the significance of the −2 log likelihood difference between models with k and $k-1$ profiles (Lo, Mendell, & Rubin, 2001). Indeed, the two-profile model was significantly better than the one-profile model ($p < .001$), the three-profile model was significantly better than the two-profile model ($p < .001$), and the four-profile model was significantly better than the three-profile model ($p < .05$). However, the five-profile model was not significantly better than the four-profile model ($p = .18$), and the six-profile model was not significantly better than the five-profile model ($p = .91$). Figure 1 displays the mean levels and standard errors of disruptive behavior at each time point for each of the four profiles.

An examination of the four-profile model indicated that this model yielded unique information and had an acceptable number of members in each profile. For profiles with a small number of members, there could be danger of a local spurious solution (Hipp & Bauer, 2006). For this model, the smallest profile represented 8% of the sample and had similar residual variances to all the other profiles. In addition, the average posterior probabilities of membership ranged from .85 to .93 across the profiles, reflecting a high degree of confidence in profile assignment. In comparison, the five- and six-profile models had profiles that represented less than 8% of the sample, with posterior probabilities below .80 (the lowest was .71).

Finally, the profiles from the four-profile model were examined for outliers and normality in a post hoc analysis. The probabilities for profile membership and most likely membership itself were saved from the analysis, and diagnostic statistics were performed on each profile. Examination of histograms, skewness, and kurtosis indicated that measures of disruptive behavior at 2, 4, and 5 years of age were normally distributed, within each profile. Examination of box plots indicated that there were no consistent outliers across the measures of disruptive behavior within the profiles. Additionally, of the individuals with a high probability of membership in each profile, on average, 55% of them were female, and gender was not significantly related to membership in any of the profiles, $\chi^2(3, 318) = 6.23, p = .10$. It is important to note that although profile membership was forced and used to perform these post hoc analyses, all other

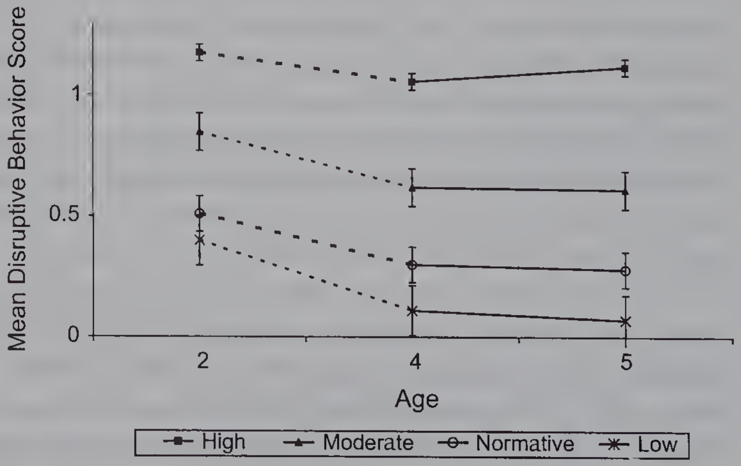


Figure 1. Profiles of disruptive behavior from 2 to 5 years of age. *Note.* Dotted lines represent the change in measurement between 2 and 4 years of age.

analyses described below were conducted within the modeling framework and in relation to the probabilities of membership in the profiles.

Description of Longitudinal Profiles

The highest profile displayed high-average levels of disruptive behavior at age 2 ($M = 1.18$, $SD = 0.09$), age 4 ($M = 1.06$, $SD = 0.04$), and age 5 ($M = 1.12$, $SD = 0.08$), and 8% of the sample had a higher probability of membership in this profile than the other three profiles. On average, this profile was just below the borderline clinical cutoff at age 2 (1.20 – 1.46; Achenbach, 1992) but by ages 4 and 5 was above the borderline clinical cutoff for boys (0.95 – 1.10; Achenbach, 1991a) and girls (0.90 – 1.00; Achenbach, 1991a); therefore, it was named the "high" profile. The second highest profile displayed average levels of disruptive behavior below the borderline clinical cutoffs, but higher than the two lower profiles, at age 2 ($M = 0.85$, $SD = 0.11$), age 4 ($M = 0.62$, $SD = 0.02$), and age 5 ($M = 0.61$, $SD = 0.04$), and 34% of the sample had a higher probability of membership in this profile than the other three profiles. Although this profile was not within the clinical range of scores, it still evidenced average elevated levels at each time point compared to the lower profiles; therefore, it was named the "moderate" profile.

Of the two lowest profiles, the second to lowest one displayed some disruptive behavior at age 2 ($M = 0.51$, $SD = 0.07$), age 4 ($M = 0.30$, $SD = 0.02$), and age 5 ($M = 0.28$, $SD = 0.02$). However, these average levels of disruptive behavior are lower than those of the profiles above. In addition, 44% of the sample had a higher probability of membership in this profile than the other three profiles. Because this profile displayed somewhat typical, but low, levels of disruptive behavior and the largest amount of the sample had a high probability of membership, it was named the "normative" profile. The lowest profile displayed low average levels of disruptive behavior at age 2 ($M = 0.40$, $SD = 0.10$), age 4 ($M = 0.11$, $SD = 0.01$), and age 5 ($M = 0.07$, $SD = 0.00$), and 14% of the sample had a higher probability of membership in this profile than the other three profiles. This profile was named the "low profile."

Prediction of Probability of Membership in Profiles

Predictors of the probabilities of membership in the above profiles were entered into the model in a hierarchical fashion (McArdle, 2005). The log likelihoods of each model were compared with the model a step before to determine whether the predictors significantly affected the model. If they did not affect the model, they were taken out to conserve power needed to detect interaction effects. First, SES was entered into the model as a covariate; however, comparison of the log likelihoods determined that it did not significantly impact the model, $\chi^2(3) = 2.0$, $p > .05$. In addition, SES did not predict the probability of membership in any of the profiles; therefore, it was removed from all further analyses. Second, the other predictors were entered in the model (frustration reactivity, physiological regulation, and maternal control). Comparison of the log likelihoods determined that they significantly impacted the model, $\chi^2(9) = 24.80$, $p < .01$. In addition, frustration reactivity and maternal control significantly predicted the probability of membership in the profiles (see Table 2). The next model included all the predictor variables and two-way interactions between the predictor variables. Interaction terms were calculated by multiplying the centered variables together. The comparison of log likelihoods between the model that included interaction terms and the model that did not determined that the two-way interactions significantly impacted the model above and beyond the main effects, $\chi^2(9) = 21.76$, $p < .01$. In addition, the interactions significantly predicted the probability of membership in the profiles (see Tables 3 & 4). Finally, a three-way interaction was tested in the model, but it did not significantly affect the model or predict membership in any of the profiles above and beyond the two-way interaction effects, $\chi^2(3) = 4.26$, $p > .05$.

There were no significant main effects that were not superseded by the interactions, and there were no significant effects differentiating the probability of

Table 2
Odds Ratios for Significant Main Effects: Low Profile as Comparison

Measure	β	*SE*	*z*	Odds
High profile				
Observed reactivity	2.51	0.71	3.52***	12.28 (0.08)
Physiological regulation	−0.45	0.31	−1.45	0.64
Maternal control	−0.22	0.19	−1.14	0.80
Moderate profile				
Observed reactivity	1.95	0.66	2.97**	7.04 (0.14)
Physiological regulation	−0.28	0.20	−1.38	0.76
Maternal control	−0.40	0.15	−2.68**	0.67 (1.49)
Normative profile				
Observed reactivity	2.35	0.67	3.51***	10.48 (0.10)
Physiological regulation	−0.21	0.21	−1.01	0.81
Maternal control	−0.46	0.17	−2.74**	0.63 (1.59)

Note. Odds ratios in parentheses are the reciprocals and refer to the odds of membership in the low profile.
$^{**}p < .01$. $^{***}p < .001$.

Table 3

Odds Ratios for Significant Interactive Effects: High Profile as Comparison

Measure	β	*SE*	*z*	Odds
Moderate profile				
Observed reactivity	−0.20	0.73	−0.27	0.82
Physiological regulation	0.33	0.29	1.15	1.39
Maternal control	−0.11	0.17	−0.61	0.89
Maternal Control × Observed Reactivity	−0.98	0.41	−2.42*	0.37 (2.70)
Maternal Control × Physiological Regulation	−0.40	0.18	−2.18*	0.67 (1.50)
Observed Reactivity × Physiological Regulation	−0.10	0.84	−0.12	0.91
Normative profile				
Observed reactivity	0.13	0.64	0.21	1.14
Physiological regulation	0.32	0.26	1.21	1.37
Maternal control	−0.18	0.17	−1.06	0.83
Maternal Control × Observed Reactivity	−0.54	0.35	−1.55	0.58
Maternal Control × Physiological Regulation	−0.33	0.18	−1.88	0.72
Observed Reactivity × Physiological Regulation	1.35	0.60	2.28*	3.87 (0.26)
Low profile				
Observed reactivity	−0.97	1.17	−0.83	0.38
Physiological regulation	0.57	0.41	1.37	1.76
Maternal control	−0.17	0.23	0.72	1.18
Maternal Control × Observed Reactivity	−1.39	0.66	−2.12*	0.25 (4.00)
Maternal Control × Physiological Regulation	−0.29	0.23	−1.30	0.75
Observed Reactivity × Physiological Regulation	−1.12	1.06	−1.05	0.33

Note. Odds ratios in parentheses are the reciprocals and refer to the odds of membership in the high profile.
$^{*}p < .05$.

membership in the moderate profile as compared to the normative profile. Therefore, only the two-way interactions are discussed and interpreted below. Table 2 depicts the significant effects found in the main effects model (found only between the low profile and other three profiles). Tables 3 and 4 present the significant effects found in the two-way interaction model, in terms of odds and corresponding significance tests, quantified as a *z* for each predictor.

Table 4

Odds Ratios With Significant Interactive Effects: Low Profile as Comparison

Measure	β	*SE*	*z*	Odds
Moderate profile				
Observed reactivity	0.77	1.10	0.70	2.16
Physiological regulation	−0.24	0.34	−0.69	0.79
Maternal control	−0.27	0.18	−1.54	0.76
Maternal Control × Observed Reactivity	0.41	0.61	0.67	1.50
Maternal Control × Physiological Regulation	−0.10	0.17	−0.59	0.90
Observed Reactivity × Physiological Regulation	1.02	1.13	0.90	2.78
Normative profile				
Observed reactivity	1.10	1.18	0.93	2.99
Physiological regulation	−0.25	0.39	−0.65	0.78
Maternal control	−0.35	0.21	−1.62	0.71
Maternal Control × Observed Reactivity	0.84	0.71	1.19	2.33
Maternal Control × Physiological Regulation	−0.04	0.23	−0.15	0.97
Observed Reactivity × Physiological Regulation	2.47	1.05	2.37*	11.87 (0.08)

Note. Odds ratios in parentheses are the reciprocals and refer to the odds of membership in the low profile.
$^{*}p < .05$.

The significant two-way interactions were tested, plotted, and interpreted in accordance with Aiken and West (1991; see also Cohen, Cohen, West, & Aiken, 2003). First, the significance of each interaction was determined through the LPA (see Tables 3 and 4), where a *z* represents the test statistic for each predictor's (i.e., interaction) effect on the probability of membership in the various profiles (similar to a *t* statistic for each β in a linear regression analysis). Second, in order to plot each interaction, high and low values (±1 *SD*) as well as the mean value (0) of one predictor (the "moderator") were entered into logistic regression equations ($Y = \beta_1 X + \beta_2 Z + \beta_3 XZ + \beta_0$) to generate three regression equations denoting the effect of the other predictor (X) on the probability of membership in the profiles (Y) at each specific value of the moderator (Z). This method is similar to what is explained in Aiken and West except that in the present case the estimates used (βs) from the LPA represented the log odds associated with the probability of membership in the various profiles. The resulting plots of these values are displayed in Figures 2 and 3. Finally, in order to fully interpret the interactions, follow-up analyses were performed in line with Aiken and West to determine what range of the moderator (Z) significantly affected (i.e., was different from 0) the relation between the other predictor (X) and the probability of membership in the profile (Y). Thus, we first explored these differences around 0 and ±1 *SD*. For most of the interactions, the relation of X and Y was significantly different from 0 at these ranges of Z. However, for the interaction between observed reactivity and maternal control predicting the probability of membership in the high versus moderate profiles, these values were not extreme enough to display these effects. Therefore, we expanded our parameters by exploring the effect around ±2 *SD* of maternal control (Z), and we found that in this range, the relation between observed reactivity (X) and the probability of membership in the high profile (Y) was significantly different from 0 around + 2 *SD* of maternal control. This suggests that maternal control needed to be more extreme to affect the probability of membership in these profiles. Descriptive statistics revealed that these values of ±1 and ±2 *SD* of the "moderators" were within range of the measures. Overall, these techniques were used in order to more fully interpret each interaction. The significance of the simple slopes (i.e., whether they are different from 0), however, does not affect the significance of the overall interaction (Aiken & West, 1991).

High Profile Versus All Other Profiles

The first comparison examined was between the high profile and the other three profiles. Compared to the low profile, observed reactivity was positively related to the probability of membership in the high profile (Table 2). However, this finding was superseded by a significant interaction between maternal control and observed reactivity (Table 3; Figure 2b). In addition, compared to the moderate profile, interactions between maternal control and observed reactivity and physiological regulation were related to the probability of membership in the high profile (Table 3; Figures 2a and 2c). Finally, an interaction between children's observed reactivity and physiological regulation was related to the probability of membership in the high profile compared to the normative profile (Table 3; Figure 3a).

Compared to the moderate profile, maternal control interacted with observed reactivity in relation to the probability of membership in the high profile (Figure 2a). Therefore, the relation between observed

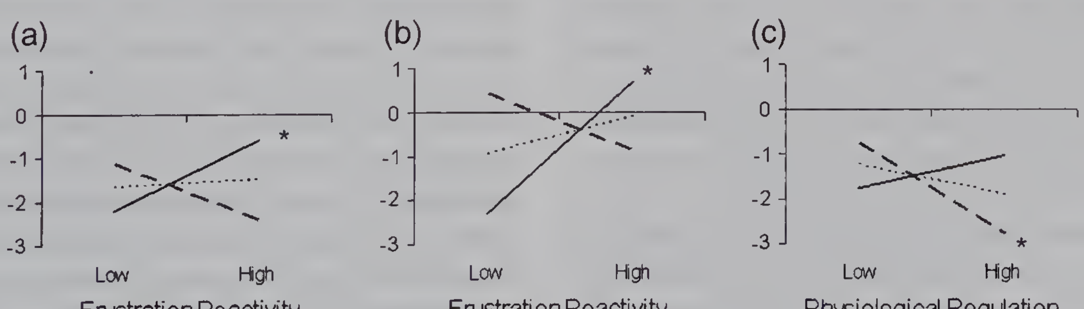


Figure 2. Interactions with maternal control (MC) predicting log odds of membership in high profile: (a) high profile versus moderate profile: MC × Frustration Reactivity, (b) high profile versus low profile: MC × Frustration Reactivity, and (c) high profile versus moderate profile: MC × Physiological Regulation.

Note. * denotes which level of the moderator showed a significant effect of MC on the log odds of membership in the high profile.

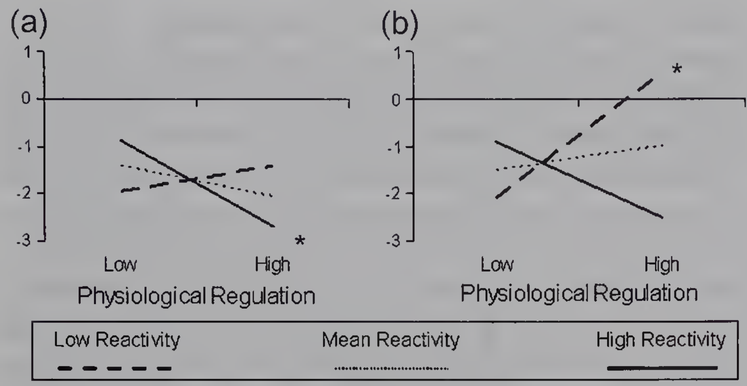


Figure 3. Interactions with frustration reactivity and physiological regulation predicting log odds of membership in high and low profiles: (a) high profile versus normative profile and (b) low profile versus normative profile.
Note. * denotes which level of the moderator showed a significant effect of frustration reactivity on the log odds of membership.

reactivity and the probability of membership in the high profile varied by the level of maternal control. For children with mothers with high control (control = 3.46), the level of observed reactivity was positively related to the probability of membership in the high profile, $z = 1.86$, $p = .06$; however, for children with mothers with low control (control = −1.73) or average control (control = 0), the effect of observed reactivity was nonsignificant, $z = -1.48$, $p = .14$. In addition, maternal control interacted with physiological regulation in relation to the probability of membership in the high profile as compared to the moderate profile (Figure 2c). Thus, the relation between physiological regulation and the probability of membership in the high profile varied by the level of maternal control. For children with mothers with high control (control = 1.73) or average control (control = 0), the effect of physiological regulation was nonsignificant, $z = 1.05$, $p = .29$; however, for children with mothers with low control (control = −1.73), the level of physiological regulation was negatively related to the probability of membership in the high profile, $z = -2.03$, $p = .04$. Overall, compared to the moderate profile, children were more likely to be in the high profile when high maternal control was combined with high observed reactivity or when low maternal control was combined with low physiological regulation.

Maternal control also interacted with observed reactivity in relation to the probability of membership in the high profile as compared to the low profile (Figure 2b). Therefore, the relation of observed reactivity to the probability of membership in the high profile varied by the level of maternal control. For children with mothers high on control (control = 1.73), the level of observed reactivity was positively related to the probability of membership in the high profile, $z = 2.44$, $p = .01$; however, the effect of observed reactivity for children with mothers low on control (control = 1.73) or average on control (control = 0) was nonsignificant, $z = -0.78$, $p = .44$. This result is similar to the interaction of maternal control and observed reactivity that differentiated the moderate profile from the high profile.

Finally, observed reactivity interacted with physiological regulation in relation to the probability of membership in the high profile as compared to the normative profile (Figure 3a). Therefore, the relation of physiological regulation to the probability of membership in the high profile varied by the level of observed reactivity. For children high on observed reactivity (reactivity = .43), the level of physiological regulation was negatively related to the probability of membership in the normative profile, $z = -2.45$, $p = .01$; however, the effect of physiological regulation for children low on observed reactivity (reactivity = −.43) or average reactivity (reactivity = 0) was nonsignificant, $z = -1.10$, $p = .27$. Consequently, children with high observed reactivity and low physiological regulation were more likely to be a member of the high profile as compared to the normative profile.

Low Profile Versus All Other Profiles

The second comparison was between the low profile and the other three profiles (Tables 2 and 4). Comparisons between the high and low profiles were discussed earlier and are not repeated. Both observed reactivity and maternal control were related to the probabilities of membership in the normative and moderate profiles in comparison to the low profile (Table 2). However, these effects were superseded by interaction effects (Table 4). Observed reactivity interacted with physiological regulation in relation to the probability of membership in the low profile compared to the normative profile (Figure 3b). Therefore, the relation of physiological regulation to the probability of membership in the low profile varied by the level of observed reactivity. For children low on observed reactivity (reactivity = −.43), the level of physiological regulation was positively related to the probability of membership in the low profile, $z = 3.25$, $p = .00$; however, the effect of physiological regulation for children high on observed reactivity (reactivity = .43) or average on observed reactivity (reactivity = 0) was nonsignificant, $z = -1.10$, $p = .27$. Overall, the probability of membership in the low profile was greater for children with low observed reactivity and high physiological regulation.

Discussion

The current investigation explored whether distinct longitudinal profiles of disruptive behavior across early childhood could be differentiated in a sample of girls and boys. In addition, observed frustration reactivity, physiological regulation, and maternal controlling behavior were examined for their joint effects on the probability of membership in the profiles. Theoretically, children's level of emotional reactivity, their ability to modulate that reactivity, and maternal behaviors that support or undermine these skills are all integral to social development (Calkins & Degnan, 2006). Furthermore, children's patterns of reactivity and regulation are thought to influence how maternal behavior is related to disruptive behavior over time.

The first goal of the study was to examine longitudinal profiles of disruptive behavior, from 2 to 5 years of age, using a semiparametric group-based statistical approach. Multiple developmental patterns of disruptive behavior were expected to emerge due to individual differences in disruptive behavior in toddlerhood and the development of behavioral regulation skills across childhood (Loeber & Stouthamer-Loeber, 1998; Moffitt, 1993). As expected, the LPA identified four profiles of disruptive behavior: a *high* profile, representing high levels of disruptive behavior reaching borderline clinical levels by age 5; a *moderate* profile, representing an elevated level of disruptive behavior at age 2 and more moderate levels at ages 4 and 5; a *normative* profile, representing moderate levels of disruptive behavior at age 2 and lower levels at ages 4 and 5; and a *low* profile, representing low levels of disruptive behavior at each age. Overall, the profiles identified in the current study support the number and composition of those trajectories of childhood externalizing behaviors found in past research (e.g., Hill et al., 2006; Shaw et al., 2003). Using a semiparametric approach to examine externalizing behavior across early childhood, there is consistent evidence for a high group, a moderate group, a low group, and a varying fourth group, with the high group typically representing 3%–14% of the sample (NICHD Early Child Care Research Network, 2004; Shaw et al., 2003). The present study identified four profiles, and 8% of the sample had a high probability of membership in the highest profile. Thus, there seems to be some consistency across studies in the patterns of disruptive behavior found in early childhood.

The second goal of this study was to explore possible predictors of these longitudinal profiles in order to further understand the etiology of developmental pathways of disruptive behavior. The joint effects of observed frustration reactivity, physiological regulation, and maternal control were examined on the probability of membership in the four profiles. Although it was hypothesized that the three predictors would all jointly relate to the probability of membership in the various profiles, the three-way interaction did not significantly impact the model or relate to the profiles. However, two-way interactions between the predictors were shown to relate to the probability of membership in the four profiles. Results revealed that maternal control interacted with observed reactivity and physiological regulation to distinguish between the probability of membership in the high profile and the moderate and low profiles. In addition, observed reactivity interacted with physiological regulation to distinguish between the probability of membership in the normative profile and the high and low profiles.

Specifically, in comparison to the high profile, the effect of observed reactivity on the probability of membership in the moderate or low profiles depended on the level of maternal control (Figures 2a and 2b). It was hypothesized that the high profile would be associated with a combination of high reactivity and low physiological regulation and higher levels of maternal control than the other profiles. As expected, children with high reactivity were more likely to be a member of the high disruptive behavior profile when their mothers were rated high on control. The level of maternal control also affected the relation of physiological regulation to the probability of membership in the moderate profile as compared to the high profile (Figure 2c). Unexpectedly, however, children with low physiological regulation were more likely to be a member of the high disruptive behavior profile when their mothers were rated low on control. Therefore, following a high profile of disruptive behavior across early childhood was related to either a combination of high maternal control and high reactivity or a combination of low maternal control and low physiological regulation. These results suggest that although maternal control may escalate child reactivity to maintain disruptive behavior problems, it may also serve as a protective factor for children with lower physiological regulation. These less regulated children might require this additional structure and direction from their mothers in order to display fewer disruptive behavior problems across early childhood. Thus, more than one constellation of maternal and child factors may increase the likelihood of having a consistently high profile of disruptive behavior across early childhood.

A second interaction between the individual child reactivity and regulation factors was found to differentiate the probability of membership in the high and normative profiles (Figure 3a). It was hypothesized that these profiles would display different combinations of the three predictors. Compared to the high profile, the normative profile was expected to show a combination of less reactivity, more regulation, and lower maternal control. As expected, for children with high observed reactivity, high physiological regulation led to a greater probability of membership in the normative profile. Thus, children's physiological regulation seemed to act as a protective factor against high levels of emotional reactivity. Whereas children who are less frustrated in general may be less likely to maintain a high disruptive behavior profile, those who are easily frustrated are at risk for doing so. Therefore, for these highly reactive children, having the physiological capacity to regulate that distress may have compensated for their temperamental nature, leading them to show lower, more normative, levels of disruptive behavior across early childhood. Without this capacity to manage their emotional reactions to frustration, these children might be more likely to display heightened levels of disruptive behavior.

Unexpectedly, the level of maternal control did not relate to any differences in the probability of membership between the high and normative profiles. Overall, maternal control may have different implications for child externalizing outcomes depending on the child's individual level of frustration reactivity and regulatory capabilities. It was hypothesized that the high profile would be associated with higher maternal control and that the normative profile would be associated with lower maternal control; however, this was not supported by the current findings. Perhaps mothers of children in the normative profile display a more moderate degree of maternal control, and this level of control may not influence the level of child disruptive behavior. In addition, for these children, maternal control may not have as profound an influence on how their reactivity or regulation contributes to their behavior problems. In fact, previous work has suggested that environmental factors have a dampened impact on children with less extreme levels of temperamental reactivity (Hane & Fox, 2007).

Observed reactivity and physiological regulation also interacted to differentiate the probability of membership in the two lower profiles (Figure 3b). It was hypothesized that these profiles would primarily differ in their level of observed frustration reactivity, with the normative profile displaying slightly more than the low profile. As expected, for children low on observed reactivity, high physiological regulation was related to having a higher probability of membership in the low profile. Therefore, children who were less reactive and more regulated were much more likely to be in the low profile than children who were more reactive or less physiologically regulated. In general, the normative profile was characterized as being less physiologically regulated and somewhat more reactive than the low profile. This result supports the notion that neither having lower emotional distress to frustrating events nor having the ability to regulate distress internally alone eliminates disruptive behavior altogether. Instead, it is the combination of both low frustration reactivity and high regulation that increases the likelihood of displaying low levels of disruptive behavior across early childhood.

Limitations

Although the present study contributes to the current literature by illuminating the complex relations among maternal behavior, frustration reactivity, emotion regulation, and longitudinal profiles of disruptive behavior, some limitations need to be addressed. Although SEMM is a useful analysis for longitudinal data, the classes (profiles) do not necessarily represent qualitatively distinct groups in the general population. Instead, they represent patterns that exist within the sample examined (Bauer & Curran, 2004). The present sample was overselected for externalizing behavior problems, and thus, the current results may not be generalizable to a randomized sample. In addition, the specific sociodemographic characteristics and measures used to describe the current sample could preclude other samples from displaying the same effects. For instance, using income level to further differentiate families with low income could be a more definitive measure of SES. The profiles identified in the current at-risk community sample will only become established with repetition and confirmation in other studies. At least one other sample of low-income boys has shown similar trajectories of conduct problems in early childhood (Shaw et al., 2003). However, future work should examine similar child and maternal factors in relation to disruptive behavior in additional samples with different demographic constellations.

Another limitation of the current study is the individual informant for the measure of disruptive behavior problems. Although the Teacher Report Form (Achenbach, 1991b) was used to confirm the level of externalizing behavior problems at 2 years of age, not every child had a day care provider at age 2, and they were not collected at 4 and 5 years of age. Similarly, the current study assessed only one example of physiological regulation to frustration. Measuring across multiple frustrating events would lend the ability to test reliability and form a more robust measure. In the future, studies should be designed to provide multiple assessments of each measure at each age point. Indeed, replication of the current findings would be enhanced if multiple informants and multiple assessments were used to test the effects of maternal and child factors on longitudinal patterns of behavior problems.

Finally, the current examination hypothesized a three-way interaction among observed frustration reactivity, physiological emotion regulation, and maternal controlling behavior. This interaction effect was not supported by the current findings. Theoretically, these aspects of the child and the environment are thought to influence each other in a transactional manner in relation to psychopathological outcomes (Calkins, 1994; Calkins & Degnan, 2006). Moreover, the two-way interactions that were observed suggest that these transactional and interactive relations most likely do exist in relation to disruptive behavior problems in early childhood. It is possible that the current sample size was not sufficient enough to capture the three-way effect above those afforded by the two-way interactions. In addition, measuring maternal behavior and child reactivity and regulation over time, and including them in the model as time-varying covariates, might help to uncover these more complex transactional relations. Finally, the measure of maternal control used in the current study may not generalize to all forms of maternal control. For instance, Bates et al. (1998) found that maternal control was related to less child reactivity when the control was aimed at decreasing dangerous behaviors that might cause harm to the child. Therefore, future studies should include measures of control in relation to play behavior and control in relation to risk-taking behavior in order to separate the possible distinguishing effects of these different forms of control on child reactivity, regulation, and disruptive behavior problems. Overall, studies should include slightly larger samples and longitudinal assessments across multiple domains in order to test these higher order interactions within a semiparametric modeling framework.

Summary and Implications

As hypothesized, four profiles of children's disruptive behavior from 2 to 5 years of age were differentiated by interactions between observed reactivity, physiological regulation, and maternal controlling behavior. Overall, results revealed that children are more likely to be in a high profile, when they display high observed reactivity combined with high maternal control or low physiological regulation combined with low maternal control. In contrast to a high profile children are more likely to be in a moderate profile when they display low observed reactivity combined with high maternal control or high physiological regulation combined with low maternal control. Moreover, children are more likely to be in a normative profile, compared to a high profile, when they display high observed reactivity combined with high physiological regulation, and children are more likely to be in a low profile, compared to a higher profile, when they display low observed reactivity combined with high maternal control or low observed reactivity combined with high physiological regulation. These findings support theory and research that posit interactive effects among emotion reactivity, regulation, and maternal behavior on externalizing behavior problems (Bates & McFayden-Ketchum, 2000; Calkins & Degnan, 2006). Furthermore, these results lend credence to a developmental psychopathological perspective, which has called for an examination of multiple developmental trajectories as well as multiple interactive child by environment processes leading to these trajectories (Kuperminc & Brookmeyer, 2006). Although the current study cannot attest to the specific mechanisms leading to these interactive processes, the findings reported here suggest that future investigations should examine the transactional mechanisms between temperament, emotion regulation, maternal behavior, and longitudinal patterns of behavior problems across early childhood.

Overall, the analysis used in the current study (i.e., LPA) improves on past methods used to identify groups. LPA allows for flexibility in profile assignment, predicts the probability of membership in a profile, and describes behavior over time even when the measurement of that behavior changes. In addition, these findings would not have been fully captured with cross-sectional or variable-oriented analyses. The differences between these methods are because LPA takes the level of disruptive behavior at each age into account, not just one time point, and allows for multiple population means rather than one mean for the whole sample. In addition, LPA accounts for missing data across the time points, allowing for increased power to detect interaction effects. In the current study, maternal control was not associated with greater disruptive behavior in general; rather, it was associated with extreme levels (high and low) of disruptive behavior across early childhood. Thus, LPA is a useful tool for examining multiple patterns of disruptive behavior during early childhood and the differentiation of those patterns with multiple indices of child and family functioning.

References

Achenbach, T. M. (1991a). *Manual for the Child Behavior Checklist/4–18 and 1991 profile*. Burlington: Department of Psychiatry, University of Vermont.

Achenbach, T. M. (1991b). *Manual for Teacher's Report Form and 1991 profile*. Burlington: Department of Psychiatry, University of Vermont.

Achenbach, T. M. (1992). *Manual for the Child Behavior Checklist/2–3 and 1992 profile*. Burlington: Department of Psychiatry, University of Vermont.

Achenbach, T. M., Edelbrock, C., & Howell, C. T. (1987). Empirically based assessment of the behavioral/emotional problems of 2- and 3-year-old children. *Journal of Abnormal Child Psychology, 15*, 629–650.

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.

Bates, J. E., & McFayden-Ketchum, S. (2000). Temperament and parent-child relations as interacting factors in children's behavioral adjustment. In V. J Molfese & D. L Molfese (Eds.), *Temperament and personality development across the life span* (pp. 141–176). Mahwah, NJ: Erlbaum.

Bates, J. E., Pettit, G. S., Dodge, K. A., & Ridge, B. (1998). Interaction of temperamental resistance to control and restrictive parenting in the development of externalizing behavior. *Developmental Psychology, 34*, 982–995.

Bauer, D. J., & Curran, P. J. (2004). The integration of continuous and discrete latent variable models: Potential problems and promising opportunities. *Psychological Methods, 9*, 3–29.

Beauchaine, T. P., Strassberg, Z., Kees, M. R., & Drabick, D. A. (2002). Cognitive response repertoires to child noncompliance by mothers of aggressive boys. *Journal of Abnormal Child Psychology, 30*, 89–101.

Belsky, J., Friedman, S. L., & Hsieh, K. H. (2001). Testing a core emotion-regulation prediction: Does early attentional persistence moderate the effect of infant negative emotionality on later development? *Child Development, 72*, 123–133.

Braungart, J. M., & Stifter, C. A. (1991). Regulation of negative reactivity during the strange situation: Temperament and attachment in 12-month-old infants. *Infant Behavior and Development, 14*, 349–364.

Broidy, L. M., Nagin, D. S., Tremblay, R. E., Bates, J. E., Brame, B., Dodge, K. A., et al. (2003). Developmental trajectories of childhood disruptive behaviors and adolescent delinquency: A six-site, cross-national study. *Developmental Psychology, 39*, 222–245.

Calkins, S. D. (1994). Origins and outcomes of individual differences in emotion regulation. In N. A. Fox (Ed.), *Emotion regulation: Behavioral and biological considerations. Monographs for the Society for Research on Child Development, 59*(2-3, Serial no. 240), 53–72.

Calkins, S. D. (1997). Cardiac vagal tone indices of temperamental reactivity and behavioral regulation in young children. *Developmental Psychobiology, 31*, 125–135.

Calkins, S. D. (2002). Does aversive behavior during toddlerhood matter? The effects of difficult temperament on maternal perceptions and behavior. *Infant Mental Health Journal, 23*, 381–402.

Calkins, S. D., & Dedmon, S. E. (2000). Physiological and behavioral regulation in two-year-old children with aggressive/destructive behavior problems. *Journal of Abnormal Child Psychology, 28*, 103–118.

Calkins, S. D., Dedmon, S. E., Gill, K. L., Lomax, L. E., & Johnson, L. M. (2002). Frustration in infancy: Implications for emotion regulation, physiological processes, and temperament. *Infancy, 3*, 175–197.

Calkins, S. D., & Degnan, K. A. (2005, April). *Temperament and adjustment in preschool: The role of physiological and behavioral response to frustration*. Poster presented at the Biennial Meeting of the Society for Research in Child Development, Atlanta, GA.

Calkins, S. D., & Degnan, K. A. (2006). Temperament in early development. In R. Ammerman (Ed.), *Comprehensive handbook of personality and psychopathology: Vol 3. Child psychopathology*. New York: Wiley.

Calkins, S. D., Gill, K. L., Johnson, M. C., & Smith, C. L. (1999). Emotional reactivity and emotional regulation strategies as predictors of social behavior with peers during toddlerhood. *Social Development, 8*, 310–334.

Calkins, S. D., Smith, C. L., Gill, K. L., & Johnson, M. C. (1998). Maternal interactive style across contexts: Relations to emotional, behavioral and physiological regulation during toddlerhood. *Social Development, 7*, 350–369.

Campbell, S. B., Pierce, E. W., Moore, G., Marakovitz, S., & Newby, K. (1996). Boys' externalizing problems at elementary school age: Pathways from early behavior problems, maternal control, and family stress. *Development and Psychopathology, 8*, 701–719.

Cicchetti, D., & Rogosch, F. A. (1996). Equifinality and multifinality in developmental psychopathology. *Development and Psychopathology, 8*, 597–600.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah, NJ: Erlbaum.

Cole, P. M., Martin, S. E., & Dennis, T. A. (2004). Emotion regulation as a scientific construct: Methodological challenges and directions for child development research. *Child Development, 75*, 317–333.

Denham, S., & Burton, R. (1996). A social-emotional intervention for at-risk 4-year-olds. *Journal of School Psychology, 34*, 225–245.

Derryberry, D., & Reed, M. A. (1994). Temperament and the self-organization of personality. *Development and Psychopathology, 6*, 653–676.

Diener, M. L., & Kim, D.-Y. (2004). Maternal and child predictors of preschool children's social competence. *Applied Developmental Psychology, 25*, 3–24.

D'Unger, A., Land, K., McCall, P., & Nagin, D. (1998). How many latent classes of delinquent/criminal careers? Results from mixed Poisson regression analyses of the London, Philadelphia, and Racine cohort studies. *American Journal of Sociology, 103*, 1593–1630.

Eisenberg, N., Fabes, R. A., & Murphy, B. C. (1996). Parents' reactions to children's negative emotions: Relations to children's social competence and comforting behavior. *Child Development, 67*, 2227–2247.

Eisenberg, N., Guthrie, I. K., Fabes, R. A., Shepard, S., Losoya, S., Murphy, B. C., et al. (2000). Prediction of elementary school children's externalizing problem behaviors from attentional and behavioral regulation and negative emotionality. *Child Development, 71*, 1367–1382.

Everitt, B., & Hand, D. J. (1981). *Finite mixture distributions*. New York: Chapman & Hall.

Fox, N. A., Henderson, H. A., Rubin, K. H., Calkins, S. D., & Schmidt, L. A. (2001). Continuity and discontinuity of behavioral inhibition and exuberance: Psychophysiological and behavioral influences across the first four years of life. *Child Development, 72*, 1–21.

Gibson, W. A. (1959). Three multivariate models: Factor analysis, latent structure analysis and latent profile analysis. *Psychometrika, 24*, 229–252.

Gilliom, M., Shaw, D. S., Beck, J. E., Schonberg, M. A., & Lukon, J. L. (2002). Anger regulation in disadvantaged preschool boys: Strategies, antecedents, and the development of self-control. *Developmental Psychology, 38*, 222–235.

Goldsmith, H. H., & Rothbart, M. K. (1993). *The Laboratory Temperament Assessment Battery: Locomotor version 2.02 (LAB-TAB)*, University of Wisconsin, Madison.

Hane, A. A., & Fox, N. A. (2007). A closer look at the transactional nature of early social development: The relations among early caregiving environments, temperament, and early social development. In F. Santoianni & C. Sabatano (Eds.), *Brain development in learning environments: Embodied and perceptual advancements* (pp. 1–15). Newcastle upon Tyne, UK: Cambridge Scholars Press.

Hill, A. L., Degnan, K. A., Calkins, S. D., & Keane, S. P. (2006). Profiles of externalizing behavior problems for boys and girls across preschool: The roles of emotion regulation and inattention. *Developmental Psychology, 42*, 913–928.

Hipp, J. R., & Bauer, D. J. (2006). Local solutions in the estimation of growth mixture models. *Psychological Methods, 11*, 36–53.

Hollingshead, A. B. (1975). *Four-factor index of social status*. New Haven, CT: Yale University.

Hubbard, J. A., Smithmyer, C. M., Ramsden, S. R., Parker, E. H., Flanagan, K. D., Dearing, K. F., et al. (2002). Observational, physiological, and self-report measures of children's anger: Relations to reactive versus proactive aggression. *Child Development, 73*, 1101–1118.

Kuperminc, G. P., & Brookmeyer, K. A. (2006). Developmental psychopathology. In R. Ammerman (Ed.), *Comprehensive handbook of personality and psychopathology: Vol. 3. Child psychopathology* (pp. 100–113). New York: Wiley.

Little, R. J., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: Wiley.

Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika, 88*, 767–778.

Loeber, R., & Stouthamer-Loeber, M. (1998). Development of juvenile aggression and violence. Some common misconceptions and controversies. *American Psychologist, 53*, 242–259.

Maughan, B., Pickles, A., Rowe, R., Costello, E. J., & Angold, A. (2000). Developmental trajectories of aggressive and non-aggressive conduct problems. *Journal of Quantitative Criminology, 16*, 199–221.

McArdle, J. J. (2005). Five steps in latent curve modeling with longitudinal life-span data. In R. Levy, P. Ghisletta, J. Le Goff, D. Spini, & E. Widmer (Eds.), *Advances in life course research: Vol. 10. Towards an interdisciplinary perspective on the life course*. Cambridge, MA: Elsevier.

Moffitt, T. E. (1993). Adolescence-limited and life-course-persistent antisocial behavior: A developmental taxonomy. *Psychological Review, 100*, 674–701.

Moore, G. A., & Calkins, S. D. (2004). Infants' vagal regulation in the still-face paradigm is related to dyadic coordination of mother-infant interaction. *Developmental Psychology, 40*, 1068–1080.

Muthén, B. (2001). Latent variable mixture modeling. In G. A Marcoulides & R. E Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 1–33). Mahwah, NJ: Erlbaum.

Muthén, B. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed.), *Handbook of quantitative methodology for the social sciences* (pp. 345–368). Newbury Park, CA: Sage.

Muthén, L., & Muthén, B. (2006). *Mplus 4.1 user's guide*. Los Angeles: Author.

Nagin, D., & Tremblay, R. E. (1999). Trajectories of boys' physical aggression, opposition, and hyperactivity on the path to physically violent and nonviolent juvenile delinquency. *Child Development, 70*, 1181–1197.

Nagin, D. S., & Tremblay, R. E. (2005). What has been learned from group-based trajectory modeling? Examples from physical aggression and other problem behaviors. *Annals of the American Academy of Political and Social Science, 602*, 82–117.

National Institute of Child Health and Human Development Early Child Care Research Network. (2004). Trajectories of physical aggression from toddlerhood to middle childhood: Predictors, correlates, and outcomes. *Monographs for the Society for Research on Child Development, 69*(4, Serial no. 278).

Owens, E. B., & Shaw, D. S. (2003). Predicting growth curves of externalizing behavior across the preschool years. *Journal of Abnormal Child Psychology, 31*, 575–590.

Paterson, G., & Sanson, A. (1999). The association of behavioural adjustment to temperament, parenting, and family characteristics among 5-year-old children. *Social Development, 8*, 293–309.

Porges, S. W. (1985). *Method and apparatus for evaluating rhythmic oscillations in aperiodic physiological response systems* (U.S. Patent no. 4520944). Washington, DC: U.S. Patent and Trademark Office.

Porges, S. W. (1991). Vagal tone: Autonomic mediator of affect. In J. A Garber & K. A. Dodge (Eds.), *The development of affect regulation and dysregulation* (pp. 111 – 128). New York: Cambridge University Press.

Porges, S. W., Doussard-Roosevelt, J. A., & Maiti, A. K. (1994). Vagal tone and the physiological regulation of emotion. *Monographs for the Society for Research on Child Development, 59*(2-3, Serial no. 240), 167 – 187.

Rothbart, M. K., Ahadi, S. A., & Evans, D. E. (2000). Temperament and personality: Origins and outcomes. *Journal of Personality and Social Psychology, 78*, 122 – 135.

Rothbart, M. K., & Bates, J. E. (2006). Temperament. In W. Damon, R. Lerner, & N. Eisenberg (Eds.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (6th ed., pp. 99 – 166). New York: Wiley.

Rothbart, M. K., Derryberry, D., & Hershey, K. L. (2000). Stability of temperament in childhood: Laboratory infant assessment to parent report at seven years. In V. J. Molfese & D. L. Molfese (Eds.), *Temperament and personality development across the life span* (pp. 85 – 119). Mahwah, NJ: Erlbaum.

Rubin, K. H., Booth, C., Rose-Krasnor, L., & Mills, R. S. L. (1995). Social relationships and social skills: A conceptual and empirical analysis. In S. Shulman (Ed.), *Close relationships and socioemotional development* (pp. 63 – 94). Norwood, NJ: Ablex.

Rubin, K. H., Burgess, K. B., Dwyer, K. M., & Hastings, P. D. (2003). Predicting preschoolers' externalizing behaviors from toddler temperament, conflict, and maternal negativity. *Developmental Psychology, 39*, 164 – 176.

Rubin, K. H., Hastings, P., Chen, X., Stewart, S., & McNichol, K. (1998). Intrapersonal and maternal correlates of aggression, conflict, and externalizing problems in toddlers. *Child Development*, *69*, 1614 – 1629.

Sanson, A., Hemphill, S. A., & Smart, D. (2004). Connections between temperament and social development: A review. *Social Development*, *13*, 142 – 170.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*, 147 – 177.

Shaw, D. S., Gilliom, M., Ingoldsby, E. M., & Nagin, D. S. (2003). Trajectories leading to school-age conduct problems. *Developmental Psychology, 39*, 189 – 200.

Shaw, D. S., Lacourse, E., & Nagin, D. S. (2005). Developmental trajectories of conduct problems and hyperactivity from ages 2 to 10. *Journal of Child Psychology and Psychiatry, 46*, 931 – 942.

Shaw, D. S., Winslow, E. B., Owens, E. B., Vondra, J. I., Cohn, J. F., & Bell, R. Q. (1998). The development of early externalizing problems among children from low-income families: A transformational perspective. *Journal of Abnormal Child Psychology, 26*, 95 – 107.

Smith, C. L., Calkins, S. D., Keane, S. P., Anastopoulos, A. D., & Shelton, T. L. (2004). Predicting stability and change in toddler behavior problems: Contributions of maternal behavior and child gender. *Developmental Psychology, 40*, 29 – 42.

Sroufe, L. A., & Rutter, M. (1984). The domain of developmental psychopathology. *Child Development*, *55*, 17 – 29.

Stifter, C. A., & Braungart, J. M. (1995). The regulation of negative reactivity in infancy: Function and development. *Developmental Psychology, 31*, 448 – 455.

Stifter, C. A., & Fox, N. A. (1990). Infant reactivity: Physiological correlates of newborn and 5-month temperament. *Developmental Psychology, 26*, 582 – 588.

Stifter, C. A., Spinrad, T. L., & Braungart-Rieker, J. M. (1999). Toward a developmental model of child compliance: The role of emotion regulation in infancy. *Child Development*, *70*, 21 – 32.

Tremblay, R. E. (2000). The development of aggressive behavior during childhood: What have we learned in the past century? *International Journal of Behavioral Development*, *24*, 129 – 141.

Winslow, E. B., Shaw, D. S., Bruns, H., & Kiebler, K. (1995, March). *Parenting as a mediator of child behavior problems and maternal stress, support, and adjustment.* Paper presented at the biennial meeting of the Society for Research in Child Development, Indianapolis, IN.

Vitaro, F., Brendgen, M., & Tremblay, R. E. (2002). Reactively and proactively aggressive children: Antecedent and subsequent characteristics. *Journal of Child Psychology and Psychiatry, 43*, 495 – 505.

Child Development, September/October 2008, Volume 79, Number 5, Pages 1377–1394

Gene–Environment Contributions to the Development of Infant Vagal Reactivity: The Interaction of Dopamine and Maternal Sensitivity

Cathi Propper
University of North Carolina–Chapel Hill

Ginger A. Moore
Pennsylvania State University

W. Roger Mills-Koonce, Carolyn Tucker Halpern, and Ashley L. Hill-Soderlund
University of North Carolina–Chapel Hill

Susan D. Calkins
University of North Carolina–Greensboro

Mary Anna Carbone
North Carolina State University

Martha Cox
University of North Carolina–Chapel Hill

This study investigated dopamine receptor genes (*DRD2* and *DRD4*) and maternal sensitivity as predictors of infant respiratory sinus arrhythmia (RSA) and RSA reactivity, purported indices of vagal tone and vagal regulation, in a challenge task at 3, 6, and 12 months in 173 infant–mother dyads. Hierarchical linear modeling (HLM) revealed that at 3 and 6 months, RSA withdrawal in response to maternal separation was greater (suggesting expected physiological regulation) in infants without the *DRD2* risk allele than those with the risk allele. At 12 months, infants with the risk allele who were also exposed to maternal sensitivity showed levels of RSA withdrawal comparable to infants who were not at genetic risk. Findings demonstrate the importance of developmental analysis of gene–environment interaction.

Respiratory sinus arrhythmia (RSA), indexing the effect of the parasympathetic nervous system on the heart, is a psychophysiological marker for the regulation of arousal, state, and reactivity underlying individual differences in emotion, behavior, and personality in children (Fox & Stifter, 1989; Porges, 1991; Stifter, 1995; Stifter, Fox, & Porges, 1989). Although there are unresolved controversies in the field regarding the measurement of vagal tone (i.e., adjusting for respiration) and how measures of RSA should be interpreted (e.g., Denver, Reed, & Porges, 2007; Grossman & Taylor, 2007), the current study will use the term RSA as an index of cardiac vagal tone and RSA reactivity to challenge as an index of vagal regulation (see Porges, 2007), acknowledging the imperfect measures of vagal functioning that these indices provide.

Due to considerable individual variation in RSA functioning (RSA and RSA reactivity) and their links to behavioral and clinical outcomes, it is critical to understand better the factors that influence their development. Although several recent studies have examined RSA and RSA reactivity as the product of environmental factors (Burgess, Marshall, Rubin, & Fox, 2003; Calkins, Smith, Gill, & Johnson, 1998; Moore & Calkins, 2004; Porter, 2003; Porter, Wouden-Miller, Silva, & Porter, 2003), there has been almost no research on potential genetic correlates, although there is research suggesting some heritability for RSA reactivity (see below). Therefore, the primary goal of this study was to explore how specific genes may interact with parenting to influence development of infant RSA reactivity. This is consistent with the recent discussion of gene–environment interactions by Moffitt, Caspi, and Rutter (2006, p. 9) that suggests testing should be theoretically driven beginning with "plausible triads" (i.e., gene, environmental factor, behavioral phenotype). Although little is known

---

This study was supported by The North Carolina Child Development Research Collaborative which is funded by the National Science Foundation through a Children's Research Initiative Grant #BCS-0126475. The authors would like to thank all the parents who participated in the Durham Child Health and Development Study and the research assistants for their valuable help in collecting this data. Also, the authors would like to thank Trudy F. C. Mackay and her laboratory at North Carolina State University for their contributions in genotyping for this project. Finally, we are grateful for the assistance and encouragement of the late Gilbert Gottlieb.

Correspondence concerning this article should be addressed to Cathi Propper, Center for Developmental Science, University of North Carolina at Chapel Hill, 100 East Franklin Street, CB# 8115, Chapel Hill, NC 27599-8115. Electronic mail may be sent to propper@email.unc.edu.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7905-0011

about candidate genes that contribute to RSA, there is a small literature linking dopamine genes and general cardiovascular function that may be relevant, a larger body of research linking specific dopamine genes with certain behavioral phenotypes (e.g., Amenta, Ricci, Rossodivita, Avola, & Tayebati, 2001; Yeh et al., 2006) that are, in turn, related to infant RSA functioning, and research linking infant RSA functioning to parenting (reviewed below). Based on this indirect evidence, we tested a theoretical relation between infants' RSA functioning (RSA and RSA reactivity) and candidate genes involved in regulating dopaminergic response (*DRD2* and *DRD4*) and the moderating effect that parenting may have on this relation.

To support the proposed theoretical model of gene–environment interaction, first we review the literature that links RSA functioning, parenting, and infant regulation related to social behavior. Next, we discuss the literature on genetics of vagal tone and cardiovascular functioning. Third, we review research linking the dopamine receptor genes *DRD2* and *DRD4* to phenotypes related to behavioral regulation.

RSA and Vagal Tone

Parasympathetic control over cardiac functioning (i.e., vagal tone) is related to self-regulation, temperament, affect, and attention (Bornstein & Suess, 2000). The term *vagal tone* refers to control of the heart via the vagus nerve and is typically measured as the amplitude of RSA (i.e., beat-to-beat heart period [HP] associated with respiration), a parasympathetic index of heart rate variability (Porges, 1996; Porges & Byrne, 1992) and underlying regulatory abilities in mammals (Porges, 1996). Baseline vagal tone is considered to be a stable neurophysiological mechanism underlying autonomic and behavioral reactivity that provides a measure of resting state in the absence of environmental challenge. In the extant literature, RSA is commonly equated with vagal tone; however, it is important to remember that this is only *one* of its components and that there are many other influences on heart rate variability (Grossman & Taylor, 2007). Of these various components, however, RSA has been the most consistently examined in relation to dimensions of behavioral functioning in infants. For example, higher resting RSA during infancy has been associated with less temperamental difficulty (Stifter & Fox, 1990), greater sustained visual attention (Richards & Cronise, 2000), secure attachment (Izard et al., 1991), more sociable and explorative behavior (Fox, 1989; Stifter et al., 1989), and greater behavioral reactivity (Porges, Doussard-Roosevelt, Portales, & Suess, 1994).

Porges's (2007) polyvagal theory of social engagement asserts that the autonomic nervous system enhances restoration and growth by regulating the "vagal brake," which slows down the heart (i.e., activated vagal tone) during situations that do not present a challenge. However, during times of environmental stress, these internal processes are disrupted and the autonomic nervous system increases metabolic output to deal with external demands. A decrease in RSA (or RSA withdrawal) typically occurs when an individual is involved in an activity that requires active coping (Porges 1991, 1996), at which time the vagal "brake" is withdrawn (i.e., vagal tone is inhibited) to support an increase in heart rate. When environmental demands have ceased, the brake is reengaged (i.e., vagal tone is activated) to promote decreases in metabolic output and a return to a calm state. Thus, effective vagal functioning has been related to the ability to maintain homeostasis in the face of changing demands by allowing a shift from attention on internal demands to external ones that include the use of coping strategies to regulate affective or behavioral arousal.

Withdrawal of RSA during a challenging situation has been related to positive outcomes in infancy such as higher soothability (Huffman et al., 1998), more attentional control (Huffman et al., 1998; Suess, Porges, & Plude, 1994), and better emotion regulation (Calkins, 1997; Porges, Doussard-Roosevelt, Portales, & Greenspan, 1996). Furthermore, individual differences in RSA functioning have been examined as an index of characteristics such as aggression (Mezzacappa et al., 1997) and hostility (Sloan et al., 1994) in adolescents and adults. Although supporting data (Bazhenova, Plonskaia, & Porges, 2001; Moore & Calkins, 2004) are not conclusive, this research has led to the assumption that greater RSA withdrawal during a challenge context reflects more effective regulation.

Influence of early parenting on RSA. Due to the almost complete reliance on caregivers in assisting with both physiological and behavioral regulation during infancy, improvement in self-organization of these processes emerges with age (Spangler & Grossmann, 1993; Spangler, Schiechle, Ilg, Maier, & Ackerman, 1994). The way in which caregivers respond to the needs of their infants may have an impact on developing adaptive methods to modulate physiological stress (Derryberry & Rothbart, 1984) by infants' progressively internalizing the regulation strategies used within the dyad during the earlier months (Calkins et al., 1998; Thompson, 1994).

RSA functioning, therefore, is a potential mediator of the relation between parenting and changes in infant emotional reactivity and regulation across development (see review by Propper & Moore, 2006). A number of studies have explored developing patterns of RSA functioning as the result of qualities of the infant–caregiver relationship and found that these qualities are, indeed, related to physiological functioning. Infants of dyads that spent more time in joint communicative sequences with continuous adjustment of behaviors in response to their partners, allowing for a range of emotional experience, had higher resting RSA than those in dyads that did not display this dyadic regulation (Porter, 2003). Moore and Calkins (2004) found that as early as 3 months of age, infants of dyads exhibiting lower synchrony showed higher physiological reactivity during a normal play episode, less RSA withdrawal during a situation meant to elicit distress, and more difficulty returning to previous levels of reactivity and RSA following distress. Yet another study found that infants of more responsive parents displayed greater regulation of heart rate (RSA was not assessed) during a challenge task than infants of less responsive parents (Haley & Stansbury, 2003).

A recent study using the same sample as in the current report examined the link between infant–mother attachment relationships and infant change in RSA during the Strange Situation Procedure (SSP; Hill et al., 2008). Findings revealed that infants classified as insecure–avoidant displayed greater cardiac arousal and RSA withdrawal during separation and reunion episodes of the procedure than did infants classified as securely attached. This finding suggests that these infants may have been experiencing high levels of internal distress while minimizing behavioral displays of their distress. These infants appeared to recruit internal resources to actively cope with an external challenge and, therefore, may have relied on self-regulation to a greater degree than other infants, who may have relied more on their mothers. These findings are consistent with earlier discussions of the attachment relationship in terms of emotion regulation strategies (Cassidy, 1994). Cassidy suggested that the attachment relationship influences, over time, the behavioral strategies used by infants to successfully interact and respond to stress within that relationship. Infants classified as insecure–avoidant may show a behavioral pattern in response to a stressful situation (i.e., separation from mother) that minimizes their emotional expressions, perhaps because they have learned from experience to rely less on the attachment relationship as a source of comfort (Belsky, Rovine, & Taylor, 1984; Main & Solomon, 1986).

Genetics of RSA. Only a handful of infant studies have examined the associations of RSA functioning among family members. Healy (1992) studied monozygotic and dizygotic twins between the ages of 11 and 35 months and found no evidence of heritability of baseline RSA. Similarly, Bornstein and Suess (2000) looked at the relation between mother and child baseline RSA at 2 months and at 5 years of age and found no concordance within dyads. They did find, however, that RSA reactivity (measured as the baseline-to-challenge task difference) was concordant between mother and child at both time points, increasing from 2 months of age ($r = .23$) to 5 years of age ($r = .42$), suggesting an increase in some shared style of physiological response to environmental challenges within the dyad over time. Through repeated interactions, the mother's behavioral approach, or affective style, may transmit to her offspring a characteristic style of autonomic responsiveness. Additionally, genetic differences may contribute to observed changes in cardiovascular regulation and may be more pronounced during conditions of stress or challenge (Boomsma, Snieder, de Geus, & van Doornen, 1998; Kupper et al., 2005).

Support for this genetic hypothesis has been found in several behavioral genetics studies. Twin studies have reported RSA heritability estimates ranging from 13% to 55% (Boomsma, van Baal, & Orlebeke, 1990; Kupper et al., 2005; Singh et al., 1999; Sinnreich, Friedlander, Luria, Sapoznikov, & Kark, 1999; Voss et al., 1996). Of note, two twin studies found an increase in genetic variance associated with RSA during a challenging situation versus RSA at rest. The first one found that roughly 50% of the variance in RSA measured in adolescence was explained by genetic factors during two challenge tasks (i.e., reaction time and mental arithmetic), whereas only 25% of the variance was explained by genetic factors during a resting period (Boomsma et al., 1990). Similarly, a study of middle-aged twins found that heritability of RSA was 35% and 43% during two comparable challenge tasks (i.e., reaction time and mental arithmetic) and decreased to approximately 31% at rest (Snieder, Boomsma, Van Doornen, & De Geus, 1997). Results of these studies support the conclusion that there is genetic influence on RSA that may be more marked during a challenge.

However, heritability estimates do not tell us anything about the specific genes that contribute to RSA functioning or in what way they do so. Only a small number of studies have examined specific genes as potential contributors to measures of cardiac functioning, most are animal models, and none of which we are aware has examined relations between specific genes and RSA. The results of the current study may provide a targeted direction for examining this relation. In addition, to date, research examining associations between specific genes and cardiovascular functioning has proceeded primarily from a health psychology approach, selecting candidate genes based on their association with disease conditions or other aspects of biological functioning. As outlined earlier, our approach was to select candidate genes based on association with behavioral phenotypes of interest. In addition, innovative prior research on the genetics of cardiovascular functioning has been conducted with samples ranging in age from adolescence to adulthood, without consideration of genetic contributions to the developing vagal system during infancy. Furthermore, in the extant literature, there is a lack of attention given to experiential factors, which may interact with genetic effects to better predict physiological outcomes (Gottlieb, 1998, 2003). The current study aimed to explore the dopamine receptor genes *DRD2* and *DRD4* that were selected on the basis of associations with behavioral phenotypes related to RSA functioning as potential predictors of RSA functioning in infancy and to examine possible interactions between these genes and infants' experience with parenting sensitivity across early infancy.

In addition to association with behavioral phenotypes, *DRD2* and *DRD4* are implicated in biological mechanisms of cardiovascular functioning. Dopamine regulates cardiovascular functioning by acting on the central and peripheral nervous systems, vascular smooth muscles, the heart, and the kidneys (Jose, Eisner, & Felder, 1999). Five distinct dopamine receptor genes have been identified and classified into two groups known as the "D1-like" super family (D1 and D5 receptors) and the "D2-like" super family (D2, D3, and D4 receptors; see Hyde, Knable, & Murray, 1996), each with its own molecular structure, mRNA coding and anatomical distribution, and chromosomal location (reviewed in Sibley & Monsma, 1992). The D2-like family of dopamine receptors has consistently been linked to the central nervous system regulation of blood pressure (Amenta et al., 2001; Bek, Eisner, Felder, & Jose, 2001; Jose, Eisner, & Felder, 1998; Yeh et al., 2006) and hypertension (Amenta, et al., 2001; Linthorst, van Giersbergen, Gras, Versteeg, & De Jong, 1994; Vaughan, van den Buuse, & Roland, 1999). There are a significant number of D2-like receptors found in the vagal complex (Hyde et al., 1996), suggesting that the *DRD2* and *DRD4* receptors are an important component of neural mechanisms regulating visceral function, including the cardiovascular system. Nevertheless, considerable research remains to be done to specify functional relations between these genes and RSA functioning in humans.

Dopamine Genes and Behavior

DRD2. Research on the associations of *DRD2* and behavioral characteristics has been limited to adolescence and adulthood, with no studies done in infancy. Variations in *DRD2* have been related to several dimensions of adult personality, including novelty seeking, sensation seeking, and aggressiveness (Cloninger, Adolfsson, & Svrakic, 1996; Noble et al., 1998). The risk allele of *DRD2* (i.e., minor *Taq*1 A1 allele) has been related to high novelty seeking scores (Noble et al., 1998; Suhara et al., 2001). Possessing the *Taq*1 A1 allele of *DRD2* has also been suggested as a risk factor for conduct disorder (Comings et al., 1996; Lu, Lee, Ko, & Lin, 2001), which is not surprising considering its association with impulsivity (Wiers, Sergeant, & Gunning, 1994) and disinhibition (McGue, Slutske, Taylor, & Iacono, 1997). Furthermore, several types of substance abuse have been associated with the *Taq*1 A1 allele of *DRD2*, including alcoholism (Blum et al., 1991; Noble et al., 1994), nicotine (Comings et al., 1996; Noble et al., 1994), and opiate dependence (Lawford et al., 2000). Finally, this allele of *DRD2* has also been associated with other addictive behaviors such as pathological gambling (e.g., Comings et al., 1996). These results are consistent with expectations, as the *Taq*1 A1 allele of *DRD2* has been reported to be associated with a decrease in D2 dopamine receptor availability (Pohjalainen et al., 1998) and the number of receptor binding sites is lowest in A1 homozygotes. Therefore, given that D2 dopamine receptors have an inhibitory function in the dopaminergic system, a lower availability of these receptors may translate into less behavioral control.

DRD4. The *DRD4* gene contains a repeated sequence polymorphism within its coding sequences that changes the length of the receptor protein that has been shown to have a moderate functional significance (Asghari et al., 1994). Individuals with longer versions of the polymorphism (L-*DRD4*; six to eight repeats) have significantly higher novelty seeking than those with the shorter version (S-*DRD4*; two to five repeats); the shorter the allele, the more efficient it is in binding dopamine (Plomin & Rutter, 1998), suggesting that the L-*DRD4* allele is a risk allele. Research on the associations of *DRD4* and behavioral characteristics has been conducted with adult and adolescent samples and more recently with infants and young children. Variations in *DRD4* have been related to dimensions of adult personality such as novelty seeking, characterized by excitability, impulsiveness, and high exploratory behavior (Benjamin et al., 1996; Ebstein, Nemanov, Klotz, Gritsenko, & Belmaker, 1997; Noble et al., 1998), cigarette smoking (Shields et al., 1998), pathological gambling (Pérez de Castro, Ibáñez, Torres, Sáiz-Ruiz, & Fernández-Piqueras, 1997), and alcoholism (George, Cheng, Nguyen, Israel, & O'Dowd, 1993). L-*DRD4* in 4-year-old children was associated with aggression as reported by mothers (Schmidt, Fox, Rubin, Hu, & Hamer, 2002). In 3-year-old children, L-*DRD4* was associated with lower intensity of emotional reactions such as happiness, anger, and sadness than S-*DRD4* (DeLuca et al., 2001). Infants at 12 months of age with the L-*DRD4* allele showed less anger-related negative emotionality than did infants with the S-*DRD4* allele in an anger-inducing task (Auerbach, Faroy, Ebstein, Kahana, & Levine, 2001). Earlier we discussed that individuals who fail to show expected RSA withdrawal in response to challenge situations may not be as reactive to those situations as most infants; infants with the L-*DRD4* may be just such an example.

In summary, the *DRD2* and *DRD4* dopamine receptor genes are associated with personality characteristics and clinical disorders that are impulsive, reward seeking, and addictive in nature. These characteristics, overall, appear to be related to a lack of regulatory ability and behavioral control, suggesting an association between dopamine and RSA, which is also highly associated with regulatory ability and behavioral control. It is important to note, however, as with most direct gene – behavior associations, many of the earlier mentioned relationships have failed to replicate (e.g., Gebhardt et al., 2000; Gelernter et al., 1997; Jönsson et al., 1998). This failure to replicate may be due to differences in sample composition, such as ethnicity or sex, or to differences in phenotype measurements. However, of interest in the current study is the possibility that it is due to the lack of research attention to environmental factors.

Goals of the Current Study

Although studies have examined the influence of parenting on the development of RSA functioning (baseline RSA and RSA reactivity to challenge), thus far, little research has examined the genetic underpinnings of this process. Research has suggested a relationship between the *DRD2* and *DRD4* genes and characteristics that are related to impulsivity and behavioral dysregulation. Similarly, studies of RSA functioning have found associations between expected RSA withdrawal in response to challenge and better behavioral regulation. Thus, the current study used a theoretically driven approach to examine the association between these genes and RSA functioning. To explore possible gene – environment mechanisms that may influence the development of RSA functioning across early infancy, this study examined the interaction between these genes and early maternal parenting sensitivity as precursors of later RSA functioning.

Method

Participants

Participants were drawn from the Durham Child Health and Development Study (DCHDS), a longitudinal sample consisting of 206 healthy, full-term infants who were followed from 3 months to 3 years of age. Families were recruited from a largely urban community via fliers and postings at birth and parenting classes, as well as through phone contact via birth records. Approximately equal numbers of European American (EA) and African American (AA) families were sampled from low- and high-income groups. Infant's race was determined by the mother (or primary caregiver); income status was assessed based on the size of the family in relation to their household income in accordance with the 2002 Federal Poverty Guidelines. Demographic information was collected during the first visit at 3 months of infant age and was updated at each subsequent visit.

Infant – mother dyads from the DCHDS were seen within a week of infants' 3-, 6-, and 12-month birthdays. Dyads that completed the mother – infant free-play task at the 6-month visit ($N = 173$) made up the initial sample for this report. However, only those dyads with infants that had complete heart rate data at *any* of the three visits, as well as complete *DRD2* genetic data, were included in analyses ($N = 142$). The final sample for analyses consisted of 67 males and 75 females. Seventy-two mothers reported their infants to be AA and 70 EA. At recruitment, the mean age of mothers was 27.8 ($SD = 5.63$, range = 18 – 40 years). Fifty-one percent of mothers reported having some college education, 49% had a high school education or some vocational training, and 44% of mothers were currently employed. Sixty-nine families reported incomes that were classified as below poverty and 73 as above poverty. Of the 142 participants with complete *DRD2* data, 11 fewer participants had complete *DRD4* data. Those with incomplete *DRD4* data did not differ by ethnicity or income.

Missing Data

Thirty-three of the dyads recruited at 3 months of infant age did not have complete data for the 6-month free-play task due to attrition ($N = 27$) or problems with coding due to video quality ($N = 6$). An additional 15 infants were missing genetic data due to difficulties with genotyping ($N = 8$) or refusal to provide a cheek cell sample ($N = 7$). Of the remaining infants, 18 were missing heart rate data at each of the three assessments. Eighty-three dyads had cardiac data at only one time point, 76 dyads had cardiac data at two time points, and 31 dyads had complete cardiac data across all three of the time points. The reasons for the missing data at each assessment, as well as the number of infants missing data per visit, are: heart rate monitor experienced problems or failure (3 months = 5, 12 months = 2), collected data contained too many artifacts to use due to movement or removal of equipment by infant (3 months = 10, 6 months = 14, 12 months = 7), infant could not finish the task due to fussiness or extreme upset (3 months = 3, 12 months = 2), or the dyad did not complete the visit or task (6 months = 4, 12 months = 7). No systematic differences associated with missing data by ethnicity or income group were found.

Procedure

Visits were conducted both at the family's home and in the lab. At each visit, infants and their mothers participated in several joint and individual tasks followed by a standardized interview and completion of demographic questionnaires by mother. Although assessments were done every 3–6 months until the child reached 36 months of age, the present study examined data only from visits that occurred when infants were 3, 6, and 12 months old when observational tasks designed to assess infants' responses to maternal separation were conducted. These visits were targeted because although the mother–infant separation tasks differed between early infancy and age 12 months, they each provided an age-appropriate standardized assessment of infants' responses to withdrawal of maternal attention, thus allowing us to compare infants' physiological regulation in response to a typical, developmentally appropriate stressor. All tasks were videotaped for later coding.

Measures

Mother–infant free play (6-month home visit). To evaluate maternal behavior during interactions with infants, mothers were provided a standard set of toys and instructed to interact with their children as they normally would if playing during some free time on a typical day. The task lasted 10 min and was videotaped for later coding.

Free-play interactions were coded by two independent coders who were unaware of the study's hypotheses. From these observations, seven subscales of maternal behavior were coded (sensitive responsiveness, intrusiveness, detachment, positive regard, negative regard, stimulation of cognitive development, and animation). Coders were trained to reliability until interclass correlation coefficients of .80 or greater were established and maintained with criterion coders (and for each individual pair of coders). All interactions were double coded and final scores were agreed on by conferencing. An overall maternal sensitivity composite was created (guided by factor analyses) by aggregating the scores for five of the subscales, including sensitivity–responsiveness, positive regard stimulation of development, animation, and detachment–disengagement (reverse scored). Similar composite scores for maternal sensitivity have been used by the National Institute of Child Health and Human Development Study of Early Child Care (1997), the Family Life Project (Blair, Granger, Willoughby, Kivlighan, & the Family Life Project Key Investigators, 2006), and other reports based on the current DCHDS sample (Mills-Koonce et al., 2007).

Mother–infant still-face procedure (3-month home visit and 6-month laboratory visit). To assess infants' responses to maternal separation, the mother–child dyad participated in the still-face procedure (SFP; Tronick, Als, Adamson, Wise, & Brazelton, 1978). Mothers placed their children in an infant seat situated on a table and then sat on a chair directly in front of them. The normal play episode consisted of mothers playing with their babies as they normally would. After this 2-min episode, mothers were told to turn away from their infants for 15 s, after which time they turned back and began the still-face episode. During the still-face episode, mothers were asked to look at their children for 2 min without any facial movements or vocalizations. Mothers were then asked to turn away from their infants for another 15 s, which was followed by a 2-min reunion episode in which mothers were told to respond to their infants in any way that they felt was appropriate. Only HP and RSA data from the still-face episode of the SFP were used for the current analyses. During this time, infants experienced an episode of withdrawal of maternal attention, which requires them to regulate their own physiological arousal without the help of their mothers. We focused specifically on this episode of the procedure because it was individual, rather than dyadic, level of regulation that was of interest in the current study and because the episode was considered most analogous to the maternal separation task at 12 months.

Strange situation procedure (12-month laboratory visit). To assess infants' responses to maternal separation at 12 months, mothers and their infants were observed in the strange situation procedure (SSP; Ainsworth, Blehar, Waters, & Wall, 1978). Throughout the procedure, infants were exposed to eight brief episodes of increasing stress, including two mother – infant separations and reunions. The current study focuses on the first episode of mother – infant separation (fourth episode) at which time mothers leave the room and the infant is alone with a stranger for 3 min. Only HP and RSA data from this first separation episode of the SSP were used for the current analyses because the second episode of separation (sixth episode) elicited high levels of distress from almost all infants, creating a possible ceiling effect.

Cardiac monitoring. At the start of each visit, the experimenter placed two disposable pediatric electrodes on the infant's chest while he or she was seated in a baby seat or mother's lap. The electrodes were connected to a preamplifier, from which the output was transmitted to a heart interbeat interval (IBI) monitor (Mini Logger; Mini-Mitter/Respironics, Bend, OR) for R-wave detection. The infant wore a smock with a large pocket in which the monitor was placed. Once the monitor was securely in place and the infant was acclimated, an electronic signal was sent to the monitor by the experimenter (by manually pressing a button on the monitor) in order to mark the start of a 2- to 4-min baseline measure of IBI activity. During this period of time, mothers were asked not to interact with, or provide toys to, their infants so that stimulation was minimized and infants' IBI could be measured accurately during a neutral and calm state. At the completion of baseline, another electronic signal was used to mark the end of this episode. IBI data were continuously collected during the rest of the visit, with electronic signals to the monitor to mark the start and end of each task. During the SFP and the SSP, electronic signals marked the start and end of each episode.

Buccal cell collection (12-month laboratory visit). DNA was obtained through the collection of infant buccal cells (i.e., cheek cells). The experimenter put on latex gloves before handling any supplies and rubbed the inside of the infants' inner cheek and gums for 20 s with a Q-tip. The Q-tip was then immediately placed into a pint-sized zip-loc bag, sealed, and put into a storage freezer where it remained until sent to the laboratory for processing. Cheek cells for DNA isolation and analysis were sent to a genetics laboratory at North Carolina State University, Raleigh, NC. All the genotyping was done blind to the study's hypotheses and outcomes.

Data Reduction

Computation of RSA and HP. At the completion of each visit, a data file containing infant IBI data for the period of collection was transferred from the monitor to a computer in the laboratory. Such activities as bodily movements, infants tugging on electrodes, physical force to the monitor, and other such disruptions may affect IBI collection by recording artifactual points within the cardiac data. Thus, the IBI data files were edited and analyzed using Mxedit software (Delta Biometrics, Bethesda, MD) by two Mxedit reliable researchers. Editing the files consisted of scanning the data for outlier points relative to adjacent data and replacing those points by dividing them or summing them so that they would be consistent with the surrounding data. Due to difficulties in collecting cardiac data from infants of this age (i.e., pulling on electrodes, equipment failure), only participants who had full and sufficient data with less than 10% editing were used in the current analyses.

The present study used Porges' (U.S. Patent No. 4,510,944, 1985) method of calculating RSA and HP. The estimate of RSA is reported in units of $\ln(\text{ms})^2$. RSA was calculated every 15 s for the baseline period and for each episode of the SFP and SSP. These epoch durations are typical for studies of short-duration tasks (Huffman et al., 1998). The mean RSA of the 15-s epochs within each episode was used in subsequent analyses. Larger values of HP indicate slower heart rate and larger values of RSA suggest greater vagal tone.

Computation of change in RSA and HP. To assess RSA reactivity, change in RSA (ΔRSA) was measured as the difference between baseline RSA and RSA during the still-face episode of the SFP at 3 and 6 months and the difference between baseline RSA and RSA during the fourth episode of the SSP at 12 months. Following previous research (Calkins, 1997; Moore & Calkins, 2004), difference scores were computed by subtracting episode RSA from baseline RSA, such that sign indicated direction of change, with positive values indicating greater RSA withdrawal (the expected response) and negative values indicating an increase in RSA during the challenge situation. Change in HP (ΔHP) was measured in the same way as ΔRSA (baseline HP minus episode HP). Positive values of HP indicated an increase in heart rate (the expected response) during the challenge situation and negative values indicated a decrease in heart rate. This method provided an index of change relative to each infant's baseline level.

Researchers have used both difference score and residual score methods to assess change in RSA. In the current data set, correlations between residual scores and difference scores for infants' RSA and HP ranged from .96 to .99. However, due to concerns that the use of difference scores might underestimate the effect between ΔRSA and infant genotype, we ran analyses using both difference scores and residualized scores and obtained the same results. Because the two scores have slightly different meanings, with residual scores indicating change if all participants started out equal in terms of baseline RSA and difference scores indicating raw change from baseline-to-challenge context, we chose the difference score method, as we believed that was most consistent with our research question.

Genotyping. Genomic DNA was extracted from each salivary sample using the Puregene DNA extraction kit by following the manufacturer's protocol for DNA isolation from 1 ml of body fluid. Saliva samples yielded DNA in adequate quantities for genotyping (approximately 200 μg/ml). Genotyping of the 48-bp repeat in Exon III of the *DRD4* gene was performed as previously described (Propper, Willoughby, Halpern, Cox, & Carbone, 2007). The genotypes s/s, s/l, or l/l were assigned to each individual where s is the short allele and l is the long allele. Based on previous results (Anchordoquy, McGeary, Liu, Krauter, & Smolen, 2003; Benjamin et al., 1996; Schmidt et al., 2002), polymorphisms made up of homogeneous short alleles (s/s) were classified as Short (S-*DRD4*) and heterogeneous polymorphisms (s/l and l/l) were classified as Long (L-*DRD4*).

Genotyping of the *DRD2* gene was performed by polymerase chain reaction amplification using the forward and reverse primers: 5′-ccgtcgacggctggccaagttgtcta-3′ (D2F1) and 5′-ccgtcgacccttcctgagtgtcatca-3′ (D2R1; Miyake et al., 1999). The amplicon was subsequently digested with the restriction enzyme, *Taq*α1 (New England Biolabs, MA). This results in digestion products of the A1 allele (310 bp) and the A2 allele (180 + 130 bp). The allele status of *DRD2* A_1^+ (A_1/A_1 and A_1/A_2 genotypes) and A_1^- (A_2/A_2 genotype) were assigned to each individual based on previous studies (see Noble, 2003).

Results

Descriptive Statistics and Bivariate Relations Among Covariates

Forty-six percent of the sample carried the risk A_1^+ allele of *DRD2*, and 36% of the sample carried the risk L-*DRD4* allele of *DRD4*. There was a significant difference in distributions of *DRD2* polymorphisms as a function of ethnicity, with AA infants more likely to have the risk A_1^+ allele of *DRD2* than EA infants, $\chi^2(1) = 3.99$, $p < .05$. This population stratification is consistent with existing research (Barr & Kidd, 1993). The impact of genetic differences by ethnicity, and the most appropriate way to address it in research designs, is currently the source of ongoing debate (Hoggart et al., 2003; Thomas & Witte, 2002; Wacholder, Rothman, & Caporaso, 2002). Because ethnicity in this study was measured by participants' self-report and individuals often are unaware of their full genetic heritage, making the measure of ethnicity subject to error, and because analyzing the groups separately would substantially attenuate power, the current analyses examined the two ethnic groups together but included ethnicity as a covariate in the model. There was no significant difference in the distribution of the *DRD4* polymorphisms as a function of ethnicity.

Maternal sensitivity was significantly higher in EA families, $t(170) = 4.08$, $p < .001$ (see Table 1) and higher income families, $t(170) = 6.10$, $p < .001$. AA infants exhibited a lower mean value of ΔRSA, reflecting less RSA withdrawal at 3 months, $t(145) = 2.23$, $p < .05$, and 6 months, $t(75) = 4.01$, $p < .05$, during the still-face episode of the SFP, and they had higher baseline RSA, on average, at 12 months, $t(99) = 3.00$, $p < .01$.

On average, infants with the *DRD2* risk allele (A_1^+) had mothers who were less sensitive than infants

Table 1
Means and Standard Deviations of Variables by Ethnicity and DRD2 *Polymorphisms*

	AA	EA		A_2/A_2 (A_1^-)	A_1/A_1 and A_1/A_2 (A_1^+)	
	M (*SD*)	*M* (*SD*)	Cohen's *d*	*M* (*SD*)	*M* (*SD*)	Cohen's *d*
Maternal sensitivity	9.38*** (4.05)	11.84*** (3.67)	.63	11.01*** (4.10)	9.65*** (4.05)	.34
Baseline RSA at 3 months	3.42 (0.95)	3.43 (1.03)		3.41 (0.92)	3.48 (1.07)	
Baseline RSA at 6 months	3.78 (0.95)	3.57 (0.80)		3.55 (0.70)	3.81 (1.08)	
Baseline RSA at 12 months	4.00** (1.00)	3.41** (0.88)	.60	3.68 (0.90)	3.65 (0.88)	
ΔRSA at 3 months	0.35* (0.90)	0.66* (0.75)	.37	0.68** (0.80)	0.26** (0.87)	.52
ΔRSA at 6 months	−0.19* (0.39)	0.39* (0.76)	.96	0.38** (0.70)	−0.08** (0.76)	.64
ΔRSA at 12 months	0.19 (0.60)	0.12 (0.70)		0.21 (0.65)	0.12 (0.64)	

Note. HP = heart period; RSA = respiratory sinus arrhythmia.
*$p < .05$. **$p < .01$. ***$p < .001$.

without the risk allele (A_1^-), $t(154) = 2.10$, $p < .05$ (see Table 1). Infants with the A_1^+ allele exhibited lower values of ΔRSA at 3 months, $t(128) = 2.92$, $p < .01$, and 6 months, $t(71) = 2.69$, $p < .01$, during the still-face episode than those without the risk allele (A_1^-). No mean differences were observed for maternal sensitivity or vagal functioning at any time point. As seen in Table 2, baseline RSA was stable over time, although ΔRSA was not. Baseline RSA was related concurrently to ΔRSA at 3 and 6 months but not at 12 months. Maternal sensitivity was uncorrelated with physiological variables.

Model Specification

HLM were estimated using SAS proc mixed in order to examine changes in children's vagal functioning (baseline RSA, ΔRSA) over three time points from 3 to 12 months. This procedure allowed for the control of the nonindependence of observations due to the same individuals being repeatedly assessed over time. An HLM approach also accounts for missing-at-random outcome data, allowing the use of all available data for the outcome of interest (Little & Rubin, 1987; Raudenbush & Bryk, 2002). Restricted maximum likelihood was used in reporting model parameters, and degrees of freedom were estimated using the Satterthwaite method. Because of the correlations between demographic factors and model parameters, ethnicity and income were included in all HLM. Models were estimated for *DRD2* and *DRD4* separately. There were no significant findings for *DRD4*, so the following reports results from the *DRD2* analyses only.

Preliminary Analyses: Vagal Functioning Over Time

Trajectories for levels of baseline RSA and levels of ΔRSA were first analyzed to assess normative change over time in these components of RSA functioning. Controlling for ethnicity and income, mean levels of baseline RSA increased from 3 to 12 months of child age by 0.25 *SD* (based on the distribution at 3 months of age), $F(2, 173) = 4.50$, $p < .05$. Mean levels of ΔRSA during the challenge task decreased from 3 to 12 months of child age, $F(2, 135) = 7.84$, $p < .001$. Because of the change in protocol to accommodate developmental changes in infants during the 1st year of life, it is not possible to determine whether changes in ΔRSA over time are due to child development or due to methodological artifact. Therefore, the following analyses of ΔRSA were conducted using *z* scores for ΔRSA rather than raw values. This approach is consistent with the main goal of the current research to analyze the effects of genetic and environmental factors on vagal functioning at different points in time as compared to analyzing their effects on trajectories of vagal functioning.

Predictors of Child Vagal Functioning Across 3 to 12 Months

Baseline RSA. There were no effects of *DRD2*, maternal sensitivity, or their interaction on baseline RSA. There was a main effect of ethnicity, $F(1, 242) = 9.46$, $p < .05$, indicating that AA infants exhibited higher levels of baseline RSA. Consistent with descriptive analyses presented earlier, this effect was moderated by time, $F(2, 135) = 3.90$, $p < .05$, such

Table 2
Correlations Among Physiological Variables

	1	2	3	4	5	6	7	8	9	10	11	12
1 Baseline HP at 3 months	—	.13	.31**	.59***	.19	.13	.61***	.26*	.13	.26**	.02	.12
2 Baseline HP at 6 months		—	.26†	−.05	−.23*	.20	.06	.50***	.27†	−.05	.12	.13
3 Baseline HP at 12 months			—	.08	−.15	.10	.22*	.30*	.73***	−.02	−.20	.20†
4 ΔHP at 3 months				—	.26*	−.03	.41***	.10	−.02	.54***	.16	.09
5 ΔHP at 6 months					—	−.18	.02	.01	−.20	.03	.50***	−.15
6 ΔHP at 12 months						—	−.03	−.03	.05	−.09	−.13	.69***
7 Baseline RSA at 3 months							—	.41***	.36***	.33***	−.11	.05
8 Baseline RSA at 6 months								—	.62***	.09	.28*	.01
9 Baseline RSA at 12 months									—	−.12	−.16	.18†
10 ΔRSA at 3 months										—	.20	.03
11 ΔRSA at 6 months											—	−.06
12 ΔRSA at 12 months												—

Note. HP = heart period; RSA = respiratory sinus arrhythmia.
$^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

that only AA infants displayed increases in baseline RSA over time, $F(2, 184) = 5.86, p < .01$.

Change in RSA. There was a significant effect of *DRD2* on ΔRSA, $F(1, 242) = 9.46, p < .01$, such that the risk A_1^+ allele was associated with lower levels of ΔRSA (i.e., less RSA withdrawal). This effect was moderated by a significant three-way interaction between *DRD2*, maternal sensitivity, and time, $F(2, 242) = 4.26, p < .05$. Probing this interaction revealed that at early ages, there were significant main effects of child genotype on ΔRSA. Children with the A_1^+ allele were observed to have lower ΔRSA, indicating less RSA withdrawal, than children without the risk allele at 3 months of age, $F(1, 107) = 6.02, p < .05$, and at 6 months of child age, $F(1, 63) = 4.23, p < .05$ (see Table 1). At 12 months of child age, a significant interaction between maternal sensitivity and child genotype was observed, $F(1, 39) = 10.02, p < .05, \eta^2 = .13$. Probing this interaction revealed that high maternal sensitivity was associated with greater ΔRSA, indicating greater RSA withdrawal for infants with the risk A_1^+ allele, $F(1, 33) = 5.29, p < .05, R^2 = .15$. As seen in Figure 1, the A_1^+ risk allele was associated with lower levels of ΔRSA at 3 and 6 months, and by 12 months, infants with highly sensitive mothers were observed to have comparable levels of ΔRSA regardless of their *DRD2* polymorphism. Cells means in Figure 1 are adjusted for covariates.

Given the exploratory nature of the current research, post hoc analyses were conducted to assess a possible confound between *DRD2* and ethnicity. First, in the model analyzed earlier, when controlling for child genotype and maternal sensitivity, the main effect of ethnicity on ΔRSA (reported in the Descriptive Statistics section) was no longer significant. Second, the interaction between sensitivity and ethnicity was examined and found to be nonsignificant, $F(1, 242) = 3.04, p < .1, \eta^2 = .06$. Figure 2 illustrates the effect of maternal sensitivity on ΔRSA for children of different genotypes, separately by ethnic groups.

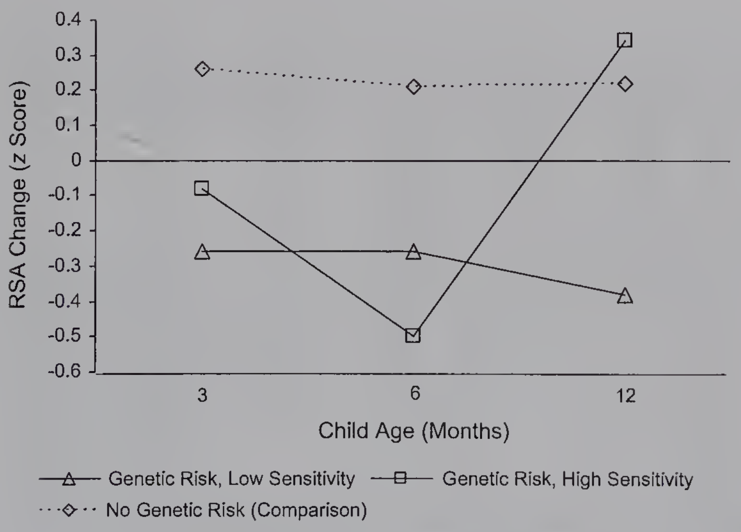


Figure 1. Maternal sensitivity moderates the relation between *DRD2* and infant change in respiratory sinus arrhythmia at 12 months.

Discussion

The primary goal of the present study was to explore genetic and environmental contributors to the development of RSA functioning related to social engagement across early infancy. The dopamine receptor genes *DRD2* and *DRD4* were selected on the basis of known associations with behavioral regulatory problems, which are, in turn, associated with infants' RSA functioning. The results of this study indicated that gene and environmental factors interacted in the development of RSA and RSA reactivity across 3–12 months, highlighting the importance of developmental analyses of gene–environment interactions (Gottlieb & Halpern, 2002).

Infants carrying the A_1^+ allele of *DRD2*, which has been identified as a risk factor for problematic outcomes in adolescents and adults related to novelty seeking, aggressiveness, and impulse control disorders, showed significantly lower levels of RSA withdrawal in reaction to a task designed to be mildly distressing at 3 and 6 months of age. As RSA withdrawal in response to challenge is a purported indicator of effective physiological regulation, these findings suggest that the *DRD2* genetic risk for behavioral disorders may be, in part, mediated by RSA functioning. Over time, however, maternal sensitivity moderated change in infant RSA reactivity to challenge. Although infants carrying the risk A_1^+ allele consistently exhibited less RSA withdrawal at 3 and 6 months, infants with the risk allele who were also exposed to sensitive maternal caregiving displayed an increase in RSA withdrawal in reaction to a distressing social situation by 12 months of age, reaching levels that were comparable to infants who were not at genetic risk.

RSA withdrawal in response to a distressing task is the typical response in infancy, and there is a body of research suggesting it indicates effective physiological regulation. Nevertheless, there are other interpretations of the current findings that cannot be ruled out. First is that infants who did not show RSA withdrawal were simply not distressed. In the current study, there is no way of accurately determining whether an infant did or did not become distressed by the procedures. Infants' behavioral and physiological indicators of distress are commonly uncorrelated in infancy (e.g., Gunnar, Mangelsdorf, Larson, &

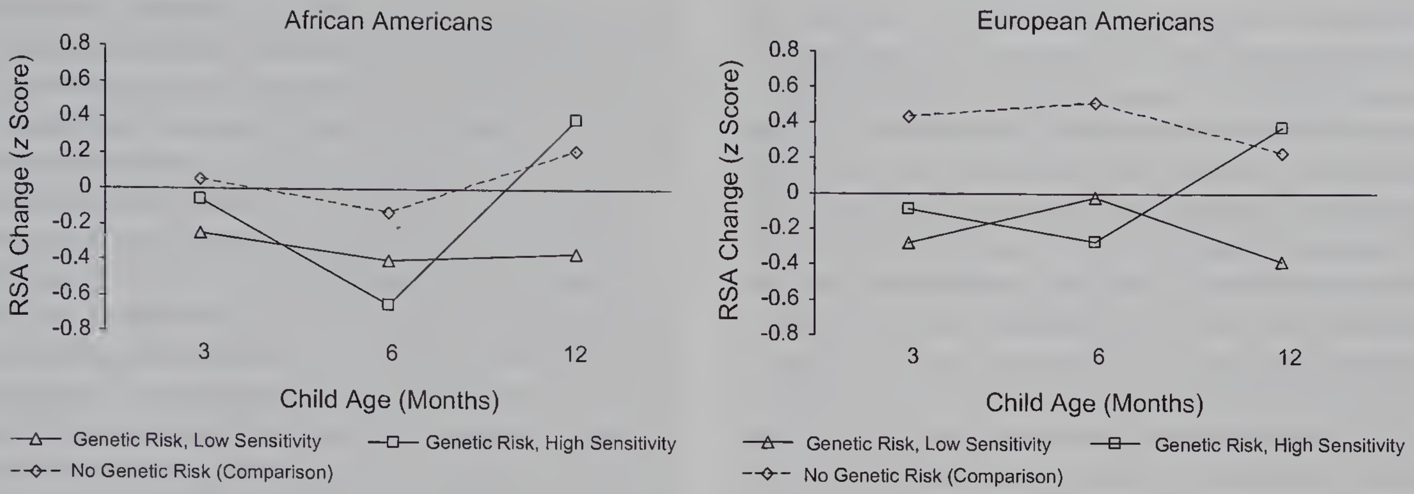


Figure 2. Interaction between maternal sensitivity and *DRD2* Risk Group by Ethnicity.

Hertsgaard, 1989; Weinberg & Tronick, 1996; cf. Bazhenova et al., 2001; Haley & Stansbury, 2003; Moore & Calkins, 2004). Measures of change in HP, which suggest arousal, were highly positively correlated with RSA reactivity, suggesting that infants who were more aroused showed greater RSA withdrawal. However, HP does not provide a measure of physiological arousal that is independent of parasympathetic influences. In future research, measures of sympathetic reactivity are needed, such as skin conductance, which has been used successfully to assess sympathetic reactivity in infants' responses (Ham & Tronick, 2006).

Second, an increase in RSA in response to a challenge task may indicate increased attention and is related to higher levels of exploratory behavior (DiPietro, Porges, & Uhly, 1992). Given that the risk A_1^+ allele of *DRD2* has been found to be associated with novelty seeking and risk-taking behavior in adults, it may be that infants with this genetic profile found the challenge tasks to be novel and interesting rather than distressing.

Third, in interpreting the interaction between genetic risk and maternal sensitivity over time, there are several possible explanations for the increase in RSA withdrawal seen for infants with the risk A_1^+ allele of *DRD2* and highly sensitive mothers. One is that genetic risk carries an associated deficit in effective vagal functioning (indicated by a lack of or diminished RSA withdrawal to an expected stressor), but as a result of experience with a sensitive mother who presumably has provided support for the development of self-regulatory skills, these infants gradually acquired more effective physiological regulation by 12 months. Alternatively, these infants may already have had adequate physiological regulation but, until they accumulated sufficient experience with a highly sensitive mother, did not become distressed at separation from her. Analogously, their counterparts in the genetic risk group who did not have highly sensitive mothers may either have had a deficit in effective vagal functioning or did not find the separation from mother distressing because of either an innate sense of novelty, as discussed earlier, or because it is not particularly distressing to be separated from a mother who is insensitive. This would be consistent with research finding that infants of depressed mothers were less likely to become distressed in response to mothers' still-face than other infants (Field, 1984).

It is also likely that, as with adults, there are individual differences in type of reactivity to stressors. Some individuals respond primarily with parasympathetic activation, some with primarily sympathetic activation, and some with activation of both systems. As stated earlier, additional measures of sympathetic reactivity are needed in future research to help to identify which of these many possible interpretations are most valid, and much larger sample sizes are needed to determine whether individual differences in patterns of autonomic reactivity can be identified in early infancy.

Although our theoretical model was based, in part, on associations between behavioral phenotypes and *DRD2* and *DRD4* genes, we found an association between infant RSA functioning and the *DRD2* gene only. This could be a sample specific finding. Additional research needs to be done to replicate this finding as well as to understand better the biological basis of the functional relation between *DRD2* and RSA functioning. Although a detailed discussion of the biological structure and function of dopamine receptors is beyond the scope of this article (see Missale, Nash, Robinson, Jaber, & Caron, 1998, for review), extant findings on the relationship between the dopaminergic system and cardiovascular functioning, mainly from animal models, may help explain the current findings. The "D2-like family" of dopamine receptors, which includes the *DRD2* and *DRD4* genes examined in the current study, has been found in the human vagal complex (Hyde et al., 1996), and dopamine receptors have been found in regions of the brain that are known to control cardiovascular function (van den Buuse, 1998), suggesting that they are an integral part of neural mechanisms that regulate cardiovascular function. However, the *DRD2* gene specifically has been associated with hypertension in both humans (Thomas, Tomlinson, & Critchley, 2000) and rats (Linthorst, De Lang, De Jong, & Versteeg, 1991; Linthorst et al., 1994). Furthermore, Li et al. (1998) found hypertension in mice to be caused by increased activity of the adrenergic nervous system after disruption to the *DRD2* gene (as cited in Bek et al., 2001), which suggests an indirect effect of dopamine on cardiac functioning. Injection of dopamine into the nucleus ambiguus of mice caused bradycardia (i.e., abnormally slow heart beat) due to the excitation of *DRD2* receptors on vagal inhibitory neurons controlling heart rate (Chitravanshi & Calaresu, 1992). Taken together, this research suggests that the relation between *DRD2* and RSA functioning found in the current study may reflect functional distinctions between the *DRD2* and *DRD4* receptor genes' effects on the vagal system. Our findings may also suggest that the relation between *DRD2* and behavioral regulatory problems is mediated by RSA functioning, whereas the relation between *DRD4* and behavioral regulatory problems is not or is related indirectly to RSA functioning.

Several methodological issues set an important context for interpreting the results. First, expected difficulties associated with obtaining valid cardiac data from young infants at three different time points resulted in a relatively small sample size. Second, because this was a longitudinal study, the challenge tasks in which we assessed infants' change in RSA needed to be developmentally appropriate. The challenge selected—maternal separation – is known to elicit stress responses from infants at various ages, thus remained consistent across age in the study. The context in which maternal separation occurred, however, had to change to accommodate developmental changes in the degree of separation that infants would find stressful. Although the still-face paradigm and the SSP are different procedures, each presents the infant with an age-appropriate disruption in social interaction that involves a removal of maternal attention, and there is evidence to show that infants' responses to the still-face paradigm predict their responses to the SSP (e.g., Cohn, Campbell, & Ross, 1991), suggesting some functional equivalence between the two procedures. It is difficult to say with certainty that the task used at 12 months was eliciting the same emotional or physiological response from infants as the task used at 3 and 6 months, but this is a fundamental dilemma of developmental research.

Another methodological issue is the possibility that population stratification could have resulted in a confound, which can arise when the outcome of interest differs between ethnic populations for nongenetic reasons while the distribution of alleles also differs by ethnicity. In our sample, AA infants were significantly more likely to have the A_1^+ allele than EA infants, consistent with extant research on the frequency of the *DRD2* A1 allele, where reported frequencies for EAs were found to be the lowest of 16 ethnic populations (Barr & Kidd, 1993). Although the issue of population stratification is an important one to consider, the degree to which this presents a problem remains controversial (Hoggart et al., 2003; Thomas & Witte, 2002), and reviews of the literature have found that the extent of population stratification bias has been exaggerated (Cardon & Palmer, 2003) and is likely to be small (Wacholder et al., 2002). Furthermore, in this study, ethnicity was assessed by mothers' self-report, and the sample was not likely to be accurately stratified in terms of genetic ancestry, especially in admixed populations where individuals may not be completely sure of their precise ancestry (Ziv & Burchard, 2003).

Exploring the possible confound suggested that, in terms of RSA reactivity, there was not a significant differential sensitivity to parenting for children of different ethnicities after controlling for genotype. Although not conclusive, this supports the interpretation of a gene–environment interaction across ethnic groups and suggests that ethnicity was not a confound. Ethnic differences found in other recent studies of genetic effects on physiological and behavioral outcomes (Propper et al., 2007; Williams et al., 2003), however, indicate the need to examine this issue in future research using larger and ethnically homogeneous samples or new methods that are being developed for detection and correction using within-family analyses or genome-wide covariates (Risch, Burchard, Ziv, & Tang, 2002; Rosenberg, Li, Ward, & Pritchard, 2003).

Finally, because there was a dearth of empirical research on which to base a search for specific candidate genes that affect the development of RSA and RSA reactivity, the justification for examining dopamine receptor genes in this study was indirect, theoretical, and based on known associations between dopamine and specific behavioral outcomes and between specific behavioral outcomes and infants' RSA reactivity. Although not ideal, this approach has been used to initiate or narrow the search for specific genes so that subsequent, molecular linkage studies may prioritize candidate genes in a hypothesis-driven search.

Keeping in mind the limitations and caveats noted earlier, the current findings provide an important first step in understanding possible genetic and environmental influences on the development of infants' RSA functioning. Although there are various interpretations of the current findings, one possibility is that these results show how infants' interactions with their caregivers over time could provide the necessary experience to effect change on genetic expression. The observed developmental change for one group of infants may reflect the impact of life experience; over time, reactions may be attributed not only to innate physiological mechanisms but also to experiential history (Sroufe, 1995). Between 3 and 12 months, infants undergo substantial developmental change, and recent developments in disparate areas of research are converging on the conclusion that interpersonal experiences strongly shape not just behavior but neural, physiological, and physical development (e.g., Siegel, 2001).

The 1st year of life is an important time for infants to develop regulatory abilities to adaptively cope with their environment. Parents may influence the way in which infants respond to situations, behaviorally and physiologically, through their help in alleviating negative emotions, reinforcing positive ones, and structuring the environment in which infants experience emotion. In the current study, at 3 and 6 months of age, infants may have exhibited RSA responses to their environments based on the influence of specific genetic polymorphisms of *DRD2* rather than parenting. However, the contributions of *DRD2* to the expression of physiological responses to the environment during the early weeks of life are likely to be, over time, only one of the many other factors that play a role in the development of these characteristics (Auerbach et al., 1999). By 12 months of age, specific experiential factors (i.e., parenting) may exert an increasingly potent influence on infants, as they become more likely to understand, interpret, remember, and apply past experiences to novel situations (Sroufe, 1990).

Thus, when examining gene – environment interactions, it is important to take into account the cumulative nature of environmental influences. Moffitt, Caspi, and Rutter (2005, p. 476) stated that "although the effects of a pathogen measured at a single point may be weak, the cumulative effects of extended or repeated exposure to that pathogen are often strong . . . most risks derive from long-standing situations rather than acute events." The current study provides empirical support for this assertion. At 3 and 6 months of age, infants may not have had enough exposure to maternal behavior for it to exert an effect; by 12 months, however, the cumulative effect of maternal behavior may be more pronounced.

In contrast to the findings for RSA reactivity, there were no genetic effects of *DRD2* on baseline RSA or its development. This is consistent with the findings of behavioral genetics studies discussed earlier in which the genetic variance associated with RSA reactivity during a challenge task was higher than that associated with baseline RSA (Boomsma et al., 1990; Kupper et al., 2005; Snieder et al., 1997). The current finding provides further evidence that conditions of challenge may elicit stronger individual differences in vagal regulation, leading genetic differences to become more pronounced (Boomsma et al., 1998).

The current findings may help to elucidate some of the nonreplications in the literature regarding the association of *DRD2* and adult personality and psychological outcomes. Extant research has reported associations between the risk allele of *DRD2* (A_1^+) and disorders related to the inability to regulate behavior (e.g., Comings et al., 1996; Noble et al., 1994), whereas other studies have failed to find these relationships (e.g., Edenberg et al., 1998). Our findings suggest that the risk allele may indeed be related to problems with physiological regulation during a stressor (i.e., lack of RSA withdrawal). However, when combined with a positive environment, the risks of this allele may diminish over time. Although the specific mechanisms that link *DRD2* and RSA functioning cannot be determined by these exploratory analyses, we provide initial data suggesting that this link may exist. Future studies should include analyses of genes and environment over time to further clarify their joint contribution to phenotypic outcomes (e.g., Caspi et al., 2002).

Another interesting finding of the current study was that infants with the risk allele (A_1^+) had mothers who were less sensitive than those without the risk allele (A_1^-). A strictly genetic explanation suggests that infants with the A_1^+ allele may have inherited this genotype from their mothers and that lack of sensitivity is associated with behavioral regulatory problems associated with this genetic risk. Another likely explanation is a "passive" gene – environment correlation, indicating that the influence of the mother's genotype on her own behavior influences her child's experiences and development (Rutter, 1997). From this perspective, our results are consistent with what we would expect if children possessed the same risk allele as their mothers; they may be at increased risk due to both genetic and environmental susceptibility for behavioral or physiological regulatory problems. Finally, children shape and select their own experiences by the way in which they act on their environment (Plomin, DeFries, & Fulker, 1988). An infant who is genetically predisposed to exhibit less RSA withdrawal in response to challenge may have more trouble regulating behavior. Infants with behavioral problems may be more difficult to care for, leading mothers to display less sensitive behavior in response.

Future Directions

This study is the first, to our knowledge, to address the interplay of genetic and environmental factors as they contribute to the development of early RSA functioning in humans. Findings have identified a number of important directions for future research, in addition to those already noted earlier. Research in molecular genetics is moving away from single gene association studies. Future research would benefit from examining multiple candidate genes in the same model (gene–gene interactions) as well as including haplotypes (i.e., sets of single alleles or closely linked genes that tend to be inherited together). Studies of this nature would require much larger sample sizes to detect multigene, multienvironment interactions.

Furthermore, differences found in maternal sensitivity as a function of ethnicity, with AA mothers being rated as less sensitive than EA mothers, are similar to reports in the literature that AA mothers show fewer displays of physical affection and warmth toward toddlers than EA mothers (Berlin, Brooks-Gunn, Spiker, & Zaslow, 1995; Bradley, Corwyn, McAdoo, & Garcia Coll, 2001). It may be the case that cultural norms dictate different parental behaviors, as well as the way in which that behavior is interpreted (Bradley et al., 2001; Deater-Deckard, Dodge, Bates, & Pettit, 1996; McLoyd & Smith, 2002). Coding systems may not be sensitive to these cultural differences. Future research should examine mechanisms, such as socioeconomic status or coding bias, which may underlie these differential effects.

In summary, the findings of the current study are the first to identify the effects of a gene–environment interaction between parenting in infancy and a specific gene that has been implicated in difficulties in behavioral self-regulation. Our findings suggest that maternal sensitivity may moderate potential negative effects of the risk allele of the dopamine gene *DRD2* on infant RSA functioning as early as 12 months of age. Although additional research is needed, the current study provides a very exciting and innovative base for future research.

References

Ainsworth, M. D., Blehar, M. C., Waters, E., & Wall, S. (1978). *Patterns of attachment: Assessed in the strange situation and at home*. Hillsdale, NJ: Erlbaum.

Amenta, F., Ricci, A., Rossodivita, I., Avola, R., & Tayebati, S. K. (2001). The dopaminergic system in hypertension. *Clinical and Experimental Hypertension, 23*, 15–24.

Anchordoquy, H. C., McGeary, C., Liu, L., Krauter, K., & Smolen, A. (2003). Genotyping of three candidate genes after whole-genome preamplification of DNA collected from buccal cells. *Behavior Genetics, 33*, 73–78.

Asghari, V., Schoots, O., van Kats, S., Ohara, K., Jovanovic, V., & Guan, H. C. et al. (1994). Dopamine D4 receptor repeat: Analysis of different native and mutant forms. *Molecular Pharmacology, 46*, 364.

Auerbach, J. G., Faroy, M., Ebstein, R., Kahana, M., & Levine, J. (2001). The association of the dopamine D4 receptor gene (DRD4) and the serotonin transporter promotor gene (5-HTTLPR) with temperament in 12-month-old infants. *Journal of Child Psychology and Psychiatry, 6*, 777–783.

Auerbach, J. G., Geller, V., Lezer, S., Shinwell, E., Belmaker, R. H., & Levine, J., et al. (1999). Dopamine D4 receptor (D4DR) and serotonin transporter promoter (5-HTTLPR) polymorphisms in the determination of temperament in 2-month-old infants. *Molecular Psychiatry, 4*, 369–373.

Barr, C. L., & Kidd, K. K. (1993). Population frequencies of the A1 allele at the dopamine D2 receptor locus. *Biological Psychiatry, 34*, 204–209.

Bazhenova, O. V., Plonskaia, O., & Porges, S. W. (2001). Vagal reactivity and affective adjustment in infants during interaction challenges. *Child Development, 72*, 1314–1326.

Bek, M. J., Eisner, G. M., Felder, R. A., & Jose, P. A. (2001). Dopamine receptors in hypertension. *Mount Sinai Journal of Medicine, 68*, 362–369.

Belsky, J., Rovine, M., & Taylor, D. G. (1984). The Pennsylvania Infant and Family Development Project: 2. The origins of individual differences in infant-mother attachment: Maternal and infant contributions. *Child Development, 59*, 156–167.

Benjamin, J., Li, L., Patterson, C., Greenberg, B. D., Murphy, D. L., & Hamer, D. H. (1996). Population and familial association between the D4 dopamine receptor gene and measures of novelty seeking. *Nature Genetics, 12*, 81–84.

Berlin, L., Brooks-Gunn, J., Spiker, D., & Zaslow, M. J. (1995). Examining observational measures of emotional support and cognitive stimulation in black and white mothers of preschoolers. *Journal of Family Issues, 16*, 664–686.

Blair, C., Granger, D., Willoughby, M., Kivlighan, K., & the Family Life Project Key Investigators. (2006). Maternal sensitivity is related to hypothalamic-pituitary-adrenal axis stress reactivity and regulation in response to emotion challenge in 6-month-old infants. *Annals of the New York Academy of Science, 1094*, 263–267.

Blum, K., Noble, E. P., Sheridan, P. J., Finley, O., Montgomery, A., & Ritchie, T., et al. (1991). Association of the A1 allele of the D2 dopamine receptor gene with severe alcoholism. *Alcohol, 8*, 409–416.

Boomsma, D. I., Snieder, H., de Geus, E. J., & van Doornen, L. J. (1998). Heritability of blood pressure increases during mental stress. *Twin Research, 1*, 15–24.

Boomsma, D. I., van Baal, G. C., & Orlebeke, J. F. (1990). Genetic influences on respiratory sinus arrhythmia across different task conditions. *Acta Geneticae Medicae et Emellologia: Twin Research (Roma), 39*, 181–191.

Bornstein, M. H., & Suess, P. E. (2000). Physiological self-regulation and information processing in infancy: Cardiac vagal tone and habituation. *Child Development, 2*, 273–287.

Bradley, R. H., Corwyn, R. F., McAdoo, H. P., & Garcia Coll, C. (2001). The home environments of children in the United States part I: Variations by age, ethnicity, and poverty status. *Child Development, 72*, 1844–1867.

Burgess, K. B., Marshall, P. J., Rubin, K. H., & Fox, N. A. (2003). Infant attachment and temperament as predictors of subsequent externalizing problems and cardiac physiology. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 44*, 819–831.

Calkins, S. D. (1997). Cardiac vagal tone indices of temperamental reactivity and behavioral regulation in young children. *Developmental Psychobiology, 31*, 125–135.

Calkins, S. D., Smith, C. L., Gill, K., & Johnson, M. C. (1998). Maternal interactive style across contexts: Relations to emotional, behavioral and physiological regulation during toddlerhood. *Social Development*, 7, 350–369.

Cardon, L. R., & Palmer, L. J. (2003). Population stratification and spurious allelic association. *Lancet, 361*, 598–604.

Caspi, A., McClay, J., Moffitt, T. E., Mill, J., Martin, J., & Craig, I. W. et al. (2002). Role of genotype in the cycle of violence in maltreated children. *Science, 297*, 851–854.

Cassidy, J. (1994). Emotion regulation: Influences of attachment relationships. In N. A. Fox (Ed.), The development of emotion regulation: Biological and behavioral considerations. *Monographs of the Society for Research in Child Development*, *59*(2–3, Serial no. 240), 228–283.

Chitravanshi, V. C., & Calaresu, F. R. (1992). Additive effects of dopamine and 8-OH-DPAT microinjected into the nucleus ambiguus in eliciting vagal bradycardia in rats. *Journal of the Autonomic Nervous System, 41*, 121–127.

Cloninger, C. R., Adolfsson, R., & Svrakic, N. M. (1996). Mapping genes for human personality. *Nature Genetics, 12*, 3–4.

Cohn, J. F., Campbell, S. B., & Ross, S. (1991). Infant response in the still-face paradigm at 6 months predicts avoidant and secure attachment at 12 months. *Development and Psychopathology, 3*, 367–376.

Comings, D. E., Rosenthal, R. J., Lesieur, H. R., Rugle, L. J., Muhleman, D., & Chiu, C. et al. (1996). A study of the dopamine D2 receptor gene in pathological gambling. *Pharmacogenetics, 6*, 223–234.

Deater-Deckard, K., Dodge, K. A., Bates, J. E., & Pettit, G. S. (1996). Discipline among African American and European American mothers: Links to children's externalizing behaviors. *Developmental Psychology, 32*, 1065–1072.

DeLuca, A., Rizzardi, M., Torrente, I., Alessandroni, R., Salvioli, G. P., & Filograsso, N. et al. (2001). Dopamine D4 receptor (DRD4) polymorphism and adaptability trait during infancy: A longitudinal study in 1- to 5-month-old neonates. *Neurogenetics, 3*, 79–82.

Denver, J. W., Reed, S. F., & Porges, S. W. (2007). Methodological issues in the quantification of respiratory sinus arrhythmia. *Biological Psychology, 74*, 286–294.

Derryberry, D., & Rothbart, M. K. (1984). Emotion, attention, and temperament. In C. E. Izard, J. Kagan, & R. B. Zajonc (Eds.), *Emotions, cognition, and behavior* (pp. 132–166). Cambridge, UK: Cambridge University Press.

DiPietro, J. A., Porges, S. W., & Uhly, B. (1992). Reactivity and developmental competence in pre-term and full-term infants. *Developmental Psychology, 28*, 831–841.

Ebstein, R. P., Nemanov, L., Klotz, I., Gritsenko, I., & Belmaker, R. H. (1997). Additional evidence for an association between the dopamine D4 receptor (D4DR) exon III polymorphism and the human personality trait of novelty seeking. *Molecular Psychiatry, 2*, 472–477.

Edenberg, H. J., Foroud, T., Koller, D. L., Goate, A., Rice, J., & Van Eerdewegh, P. et al. (1998). A family-based analysis of the association of the dopamine D2 Receptor (*DRD2*) with alcoholism. *Alcoholism: Clinical and Experimental Research, 22*, 505–512.

Field, T. (1984). Early interactions between infants and their postpartum depressed mothers. *Infant Behavior and Development*, 7, 517–522.

Fox, N. A. (1989). Psychophysiological correlates of emotional reactivity during the first year of life. *Developmental Psychology, 25*, 364–372.

Fox, N. A., & Stifter, C. A. (1989). Biological and behavioral differences in infant reactivity and regulations. In G. Kohnstamm, J. Bates, & M. Rothbart (Eds.), *Handbook of temperament in childhood*. Chichester, UK: John Wiley.

Gebhardt, C., Leisch, F., Schussler, P., Fuchs, K., Stompe, T., & Sieghart, W., et al. (2000). Non-association of dopamine D4 and D2 receptor genes with personality in healthy individuals. *Psychiatric Genetics, 10*, 131–137.

Gelernter, J., Kranzler, H., Coccaro, E., Siever, L., New, A., & Mulgrew, C. L. (1997). D4 dopamine-receptor (DRD4) alleles and novelty seeking in substance-dependent, personality-disorder, and control subjects. *American Journal of Human Genetics*, 61, 1144–1152.

George, S. R., Cheng, R., Nguyen, T., Israel, Y., & O'Dowd, B. F. (1993). Polymorphisms of the D4 dopamine receptor alleles in chronic alcoholism. *Biochemical and Biophysical Research Communications, 196*, 107–114.

Gottlieb, G. (1998). Normally occurring environmental and behavioral influences on gene activity: From central dogma to probabilistic epigenesis. *Psychological Review, 105*, 792–802.

Gottlieb, G. (2003). On making behavioral genetics truly developmental. *Human Development, 46*, 337–355.

Gottlieb, G., & Halpern, C. T. (2002). A relational view of causality in normal and abnormal development. *Development and Psychopathology, 14*, 421–435.

Grossman, P., & Taylor, E. (2007). Toward understanding sinus arrhythmia relators to cardiac vagal tune, evolution, and biobehavioral functions. *Biological Psychology, 74*, 263–285.

Gunnar, M. R., Mangelsdorf, S., Larson, M., & Hertsgaard, L. (1989). Attachment, temperament, and adrenocortical activity in infancy: A study of psychoendocrine regulation. *Developmental Psychology, 25*, 355–363.

Haley, D. W., & Stansbury, K. (2003). Infant stress and parent responsiveness: Regulation of physiology and behavior during still-face and reunion. *Child Development, 74*, 1534–1546.

Ham, J., & Tronick, E. (2006). Infant resilience to the stress of the still-face: Infant and maternal psychophysiology are related. *Annals of the New York Academy of Sciences, 1094*, 297–302.

Healy, B. T. (1992). The heritability of autonomic nervous system processes. In T. M. Field, P. M. McCabe, & N. Schneiderman (Eds.), *Stress and coping in infancy and childhood* (pp. 69–82). Hillsdale, NJ: Erlbaum.

Hill, A., Mills-Koonce, R., Propper, C., Calkins, S., Granger, D., Moore, G., et al. (2008). *Physiological responses to the strange situation: Vagal withdrawal, salivary α-amylase and salivary cortisol in infants and mothers as a function of attachment status.* Manuscript under review.

Hoggart, C. J., Parra, E. J., Shriver, M. D., Bonilla, C., Kittles, R. A., Clayton, D. G., et al. (2003). Control of confounding of genetic associations in stratified populations. *American Journal of Human Genetics, 72*, 1492–1504.

Huffman, L. C., Bryan, Y., Del Carmen, R., Pederson, F., Doussard-Roosevelt, J., & Porges, S. W. (1998). Infant temperament and cardiac vagal tone: Assessments at twelve weeks of age. *Child Development, 69*, 624–635.

Hyde, T. M., Knable, M. B., & Murray, A. M. (1996). Distribution of dopamine D1-D4 receptor subypes in human dorsal vagal complex. *Synapse, 24*, 224–232.

Izard, C. E., Porges, S. W., Simons, R. F., Haynes, O. M., Hyde, C., & Parisi, M. et al. (1991). Infant cardiac activity: Developmental changes and relations with attachment. *Developmental Psychology, 27*, 432–439.

Jönsson, E. G., Nöthen, M. M., Gustavsson, J. P., Neidt, H., Forslund, K., Mattila-Evenden, N., et al. (1998). Lack of association between dopamine D4 receptor gene and personality traits. *Psychological Medicine, 28*, 985–989.

Jose, P. A., Eisner, G. M., & Felder, R. A. (1998). The renal dopamine receptors in health and hypertension. *Pharmacological Therapy, 80*, 149–182.

Jose, P. A., Eisner, G. M., & Felder, R. A. (1999). Dopaminergic mechanisms in the development of hypertension. In R. McCarty, D. A. Blizard, & R. I. Chevalier (Eds.), *Handbook of hypertension* (pp. 1–44). Amsterdam: Elsevier Science.

Kupper, N., Gonneke, W., Posthuma, D., de Boer, D., Boomsma, D. I., & de Geus, E. J. (2005). A genetic analysis of ambulatory cardiorespiratory coupling. *Psychophysiology, 42*, 202–212.

Lawford, B. R., Young, R. M., Noble, E. P., Sargent, J., Rowell, J., & Shadforth, S., et al. (2000). The D2 dopamine receptor A(1) allele and opioid dependence: Association with heroin use and response to methadone treatment. *American Journal of Medical Genetics, 96*, 592–598.

Li, X. X., Asico, L. D., Low, M., et al. (1998). Disruption of the dopamine D2 receptor produces non-renal dependent hypertension. *Journal of the American Society of Nephrology, 9*, 311.

Linthorst, A. C. E., De Lang, H., De Jong, W. T., & Versteeg, D. H. G. (1991). Effect of the dopamine D2 receptor agonist quinpirole on the in vivo release of dopamine in the caudate nucleus of hypertensive rats. *European Journal of Pharmacology, 201*, 125–133.

Linthorst, A. C. E., van Giersbergen, P. L., Gras, M., Versteeg, D. H. G., & De Jong, W. (1994). The nigrostriatal dopamine system: Role in the development of hypertension in spontaneously hypertensive rats. *Brain Research, 639*, 261–268.

Little, R. A. A., & Rubin, D. B. (1987). *Statistical analysis with missing data.* New York: Wiley.

Lu, R. B., Lee, J. F., Ko, H. C., & Lin, W. W. (2001). Dopamine D2 receptor gene (DRD2) is associated with alcoholism with conduct disorder. *Alcoholism: Clinical and Experimental Research, 25*, 177–184.

Main, M., & Solomon, J. (1986). Discovery of a new, insecure-disorganized/disoriented attachment pattern. In T. B. Brazelton & M. Yogman (Eds.), *Affective development in infancy* (pp. 95–124). Norwood, NJ: Ablex.

McGue, M., Slutske, W., Taylor, J., & Iacono, W. G. (1997). Personality and substance use disorders: I. Effects of gender and alcoholism subtype. *Alcoholism: Clinical and Experimental Research, 21*, 513–520.

McLoyd, V. C., & Smith, J. (2002). Physical discipline and behavior problems in African American, European American, and Hispanic children: Emotional support as a moderator. *Journal of Marriage and Family, 64*, 40–53.

Mezzacappa, E., Tremblay, R. E., Kindlon, D., Saul, J. P., Arseneault, L., & Seguin, J. et al. (1997). Anxiety, antisocial behavior, and heart rate regulation in adolescent males. *Journal of Child Psychology and Psychiatry, 38*, 457–469.

Mills-Koonce, W., Gariépy, J., Propper, C., Sutton, K., Calkins, S., Moore, G., et al. (2007). Infant and parent factors associated with early maternal sensitivity: A caregiver-attachment systems approach. *Infant Behavior and Development, 30*, 114–126.

Missale, C., Nash, S. R., Robinson, S. W., Jaber, M., & Caron, M. G. (1998). Dopamine receptors: From structure to function. *Physiological Review, 78*, 189–225.

Miyake, H., Nagashima, K., Onigata, K., Nagashima, T., Takano, Y., & Morikawa, A. (1999). Allelic variations of the D2 dopamine receptor gene in children with idiopathic short stature. *Journal of Human Genetics, 44*, 26–29.

Moffitt, T. E., Caspi, A., & Rutter, M. (2005). Strategy for investigating interactions between measured genes and measured environments. *Archives of General Psychiatry, 62*, 473–481.

Moffitt, T. E., Caspi, A., & Rutter, M. (2006). Measured gene-environment interactions in psychopathology: Concepts, research strategies, and implications for research, intervention, and public understanding of genetics. *Perspectives on Psychological Science, 1*, 5–27.

Moore, G. A., & Calkins, S. D. (2004). Infants' vagal regulation in the still-face paradigm is related to dyadic coordination of mother-infant interaction. *Developmental Psychology, 40*, 1068–1080.

National Institute of Child Health and Human Development Study of Early Child Care. (1997). Effects of infant child care on infant-mother attachment security: Results of the NICHD study of early child care. *Child Development, 68*, 860–879.

Noble, E. P. (2003). D2 dopamine receptor gene in psychiatric and neurologic disorders and its phenotypes. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 116B*, 103–125.

Noble, E. P., Ozkaragoz, T. Z., Ritchie, T. L., Zhang, X., Belin, T. R., & Sparkes, R. S. (1998). D2 and D4 dopamine receptor polymorphisms and personality. *American Journal of Medical Genetics: Neuropsychiatric Genetics, 81*, 257–267.

Noble, E. P., Syndulko, K., Fitch, R. J., Ritchie, T., Bohlman, M. C., Guth, P. et al. (1994). D2 dopamine receptor *Taq*1 A alleles in medically ill alcoholic and nonalcoholic patients. *Alcohol, 29*, 729–744.

Pérez de Castro, I., Ibáñez, A., Torres, P., Sáiz-Ruiz, J., & Fernández-Piqueras, J. (1997). Genetic association study between pathological gambling and a functional DNA polymorphism at the D4 receptor gene. *Pharmacogenetics, 7*, 345–348.

Plomin, R., DeFries, J. C., & Fulker, D. W. (1988). *Nature and nurture during infancy and early childhood*. New York: Cambridge University Press.

Plomin, R., & Rutter, M. (1998). Child development, molecular genetics, and what to do with genes once they are found. *Child Development, 69*, 1223–1242.

Pohjalainen, T., Rinne, J. O., Nagren, K., Lehikoinen, P., Anttila, K., Syvalahti, E. K. G., et al. (1998). The A1 allele of the human D2 dopamine receptor gene predicts low D2 receptor availability in healthy volunteers. *Molecular Psychiatry, 3*, 256–260.

Porges, S. W. (1985). *Method and apparatus for evaluating rhythmic oscillations in a periodic physiological response system* (U.S. Patent No. 4,510,944). Washington, DC: U.S. Patent Office.

Porges, S. W. (1991). Vagal tone: An autonomic mediatory of affect. In J. A. Garber & K. A. Dodge (Eds.), *The development of affect regulation and dsyregulation* (pp. 11–128). New York: Cambridge University Press.

Porges, S. W. (1996). Physiological regulation in high-risk infants: A model for assessment and potential intervention. *Development and Psychopathology, 8*, 43–58.

Porges, S. W. (2007). The polyvagal perspective. *Biological Psychology, 74*, 116–143.

Porges, S. W., & Byrne, E. A. (1992). Research methods for measurement of heart rate and respiration. *Biological Psychology, 34*, 93–130.

Porges, S. W., Doussard-Roosevelt, J. A., Portales, A. L., & Greenspan, S. I. (1996). Infant regulation of the vagal "brake" predicts child behavior problems: A psychobiological model of social behavior. *Developmental Psychobiology, 29*, 697–712.

Porges, S. W., Doussard-Roosevelt, J. A., Portales, A. L., & Suess, P. E. (1994). Cardiac vagal tone: Stability and relation to difficultness in infants and 3-year-olds. *Developmental Psychobiology, 27*, 289–300.

Porter, C. L. (2003). Coregulation in mother-infant dyads: Links to infants' cardiac vagal tone. *Psychological Reports, 92*, 307–319.

Porter, C. L., Wouden-Miller, M., Silva, S. S., & Porter, A. E. (2003). Marital harmony and conflict: Links to infants' emotional regulation and cardiac vagal tone. *Infancy, 4*, 297–307.

Propper, C., & Moore, G. A. (2006). The influence of parenting on infant emotionality: A multi-level psychobiological perspective. *Developmental Review, 26*, 427–460.

Propper, C., Willoughby, M., Halpern, C. T., Cox, M., & Carbone, M. A. (2007). Parenting quality, DRD4, and the prediction of externalizing and internalizing behaviors in early childhood. *Developmental Psychobiology, 49*, 619–632.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models*. Thousand Oaks, CA: Sage.

Richards, J. E., & Cronise, K. (2000). Extended visual fixation in the early preschool years: Look duration, heart rate changes, and attentional inertia. *Child Development, 71*, 602–620.

Risch, N., Burchard, E., Ziv, E., & Tang, H. (2002). Categorization of humans in biomedical research: Genes, race and disease. *Genome Biology, 3*, 1–12.

Rosenberg, N. A., Li, L. M., Ward, R., & Pritchard, J. K. (2003). Informativeness of genetic markers for inference of ancestry. *American Journal of Human Genetics, 73*, 1402–1422.

Rutter, M. L. (1997). Nature-nurture integration: The example of antisocial behavior. *American Psychologist, 52*, 390–98.

Schmidt, L. A., Fox, N. A., Rubin, K. H., Hu, S., & Hamer, D. H. (2002). Molecular genetics of shyness and aggression in preschoolers. *Personality and Individual Differences, 33*, 227–238.

Shields, P., Lerman, C., Audrain, J., Main, D., Boyd, N., & Caporaso, N. (1998). Dopamine D4 receptors and the risk of cigarette smoking in African-Americans and Caucasians. *Cancer Epidemiology, Biomarkers, and Prevention, 7*, 453–458.

Sibley, D., & Monsma, F. (1992). Molecular biology and dopamine receptors. *Trends in Pharmacological Science, 13,* 61–69.

Siegel, D. J. (2001). Toward an interpersonal neurobiology of the developing mind: Attachment relationships, "mindsight," and neural integration. *Infant Mental Health Journal, 22,* 67–94.

Singh, J. P., Larson, M. G., O'Donnell, C. J., Tsuji, H., Evans, J. C., & Levy, D. (1999). Heritability of heart rate variability: The Framingham Heart Study. *Circulation, 99,* 2251–2254.

Sinnreich, R., Friedlander, Y., Luria, M. H., Sapoznikov, D., & Kark, J. D. (1999). Inheritance of heart rate variability: The kibbutzim family study. *Human Genetics, 105,* 654–661.

Sloan, R. P., Shapiro, P. A., Bigger, J. T., Bagiella, M., Steinman, R. C., & Gorman, J. M. (1994). Cardiac autonomic control and hostility in healthy subjects. *American Journal of Cardiology, 74,* 298–300.

Snieder, H., Boomsma, D. I., Van Doornen, L. J., & De Geus, E. J. (1997). Heritability of respiratory sinus arrhythmia: Dependency on task and respiration rate. *Psychophysiology, 34,* 317–328.

Spangler, G., & Grossmann, K. E. (1993). Biobehavioral organization in securely and insecurely attached infants. *Child Development, 64,* 1439–1450.

Spangler, G., Schiechle, M., Ilg, U., Maier, U., & Ackerman, C. (1994). Maternal sensitivity as an external organizer for biobehavioral regulation in infancy. *Developmental Psychobiology, 27,* 425–437.

Sroufe, L. A. (1990). An organizational perspective on the self. In D. Cicchetti & M. Beeghly (Eds.), *The self in transition: Infancy to childhood* (pp. 281–307). Chicago: University of Chicago Press.

Sroufe, L. A. (1995). *Emotional development: The organization of emotional life in the early years.* Cambridge, UK: Cambridge University Press.

Stifter, C. A. (1995). Approach/withdrawal processes in infancy: The relationship between parasympathetic tone and infant temperament. In K. Hood, E. Tobach, & G. Greenberg (Eds.), *Approach/withdrawal theory and behavioral development* (pp. 371–399). Hillsdale, NJ: Erlbaum.

Stifter, C. A., & Fox, N. A. (1990). Infant reactivity: Physiological correlates of newborn and 5-month temperament. *Developmental Psychology, 26,* 582–588.

Stifter, C. A., Fox, N. A., & Porges, S. W. (1989). Facial expressivity and vagal tone in 5- and 10-month-old infants. *Infant Behavior and Development, 12,* 127–137.

Suess, P. E., Porges, S. W., & Plude, D. J. (1994). Cardiac vagal tone and sustained attention in school-age children. *Psychophysiology, 31,* 17–22.

Suhara, T., Yasuno, F., Sudo, Y., Yamamoto, M., Inoue, M., Okubo, Y., et al. (2001). Dopamine D_2 receptors in the insular cortex and the personality trait of novelty seeking. *Neuroimage, 13,* 891–895.

Thomas, G. N., Tomlinson, B., & Critchley, J. A. J. H. (2000). Modulation of blood pressure and obesity with the dopamine D2 receptor gene *Taq*1 polymorphism. *Hypertension, 36,* 177–182.

Thomas, D. C., & Witte, J. S. (2002) Point: Population stratification: A problem for case-control studies of candidate-gene associations? *Cancer Epidemiology, Biomarkers and Prevention, 11,* 505–512.

Thompson, R. A. (1994). Emotion regulation: A theme in search of a definition. In N. A. Fox (Ed.), Emotion regulation: Behavioral and biological considerations. *Monographs of the Society for Research in Child Development* (2–3, Serial no. 240), 25–52.

Tronick, E. Z., Als, H., Adamson, L., Wise, S., & Brazelton, B. (1978). The infant's response to entrapment between contradictory messages in face-to-face interaction. *Journal of the American Academy of Child Psychiatry, 1,* 1–13.

van den Buuse, M. (1998). Role of the mesolimbic dopamine system in cardiovascular homeostasis. Stimulation of the ventral tegmental area modulates the effect of vasopressin on blood pressure in conscious rats. *Clinical and Experimental Pharmacology and Physiology, 25,* 661–668.

Vaughan, C. E., van den Buuse, M., & Roland, B. L. (1999). Brain dopamine D2 receptor mRNA levels are elevated in young spontaneously hypertensive rats. *Neuroscience Research, 34,* 199–205.

Voss, A., Busjahn, A., Wessel, N., Schurath, R., Faulhaber, H. D., Luft, F. C., et al. (1996). Familial and genetic influences on heart rate variability. *Journal of Electrocardiology, 29,* 154–160.

Wacholder, S., Rothman, N. & Caporaso, N. (2002). Counterpoint: Bias from population stratification is not a major threat to the validity of conclusions from epidemiological studies of common polymorphisms and cancer. *Cancer Epidemiology, Biomarkers and Prevention, 11,* 513–520.

Weinberg, M. K., & Tronick, E. Z. (1996). Infant affective reactions to the resumption of maternal interaction after the still-face. *Child Development, 67,* 905–914.

Wiers, R. W., Sergeant, J. A., & Gunning, W. B. (1994) Psychological mechanisms of enhanced risk of addiction in children of alcoholics: A dual pathway? *Acta Paediatrica Supplement, 404,* 9–13.

Williams, R. B., Marchuk, D. A., Gadde, K. M., Barefoot, J. C., Crichnik, K., Helms, M. J., et al. (2003). Serotonin-related gene polymorphisms and central nervous system 2serotonin function. *Neuropsychopharmacology, 28,* 533–541.

Yeh, T., Yang, Y., Chiu, N., Yao, W., Yeh, S., Wu, J., et al. (2006). Correlation between striatal dopamine D2/D3 receptor binding and cardiovascular activity in healthy subjects. *American Journal of Hypertension, 19,* 964–969.

Ziv, E., & Burchard, E. G. (2003). Human population structure and genetic association studies. *Pharmacogenomics, 4,* 431–441.

Child Development, September/October 2008, Volume 79, Number 5, Pages 1395–1414

The Family Check-Up With High-Risk Indigent Families: Preventing Problem Behavior by Increasing Parents' Positive Behavior Support in Early Childhood

Thomas J. Dishion
University of Oregon

Daniel Shaw
University of Pittsburgh

Arin Connell
Case Western Reserve University

Frances Gardner
Oxford University

Chelsea Weaver
University of Pittsburgh

Melvin Wilson
University of Virginia

Seven hundred thirty-one income-eligible families in 3 geographical regions who were enrolled in a national food supplement program were screened and randomized to a brief family intervention. At child ages 2 and 3, the intervention group caregivers were offered the Family Check-Up and linked parenting support services. Latent growth models on caregiver reports at child ages 2, 3, and 4 revealed decreased behavior problems when compared with the control group. Intervention effects occurred predominantly among families reporting high levels of problem behavior at child age 2. Families in the intervention condition improved on direct observation measures of caregivers' positive behavior support at child ages 2 and 3; improvements in positive behavior support mediated improvements in children's early problem behavior.

Problem behavior in middle childhood and adolescence is a predictable outgrowth of problematic adjustment in early childhood (Brook, Whiteman, Cohen, & Tanaka, 1992; Brook, Whiteman, Gordon, & Cohen, 1986; Campbell, Shaw, & Gilliom, 2000; Caspi, Moffitt, Newman, & Silva, 1998; Hawkins, Lishner, Catalano, & Howard, 1986; Lyons-Ruth, Zeanah, & Benoit, 1996; Shaw, Gilliom, Ingoldsby, & Nagin, 2003; Vicary & Lerner, 1983). Longitudinal studies begun with children as young as age 3 years (e.g., Caspi et al., 1998) have revealed an association between early behavior problems and long-term profiles of risk, including substance dependence in young adulthood.

Our study was motivated by three principles: (a) the prevention of problem behavior in children and adolescents must be family centered and ecologically focused; (b) for the most effective long-term impact on health, empirically supported family interventions must be brief and therefore more cost effective and must be embedded in existing service systems (Hoagwood & Koretz, 1996); and (c) developmental transition points such as toddlerhood, school entry, and early adolescence offer unique opportunities for health promotion and risk reduction because child and family behaviors reorganize at these points (Dishion & Stormshak, 2007; Sameroff & Fiese, 1987).

Intervening during the first of these transition points, when the child begins to walk and becomes physically autonomous (ages 1–2 years), is an important strategy for preventing adolescent problem behaviors such as delinquency and substance use. Parent–child interaction patterns during this transitional developmental period are challenged by normative increases in child mobility, language, independence, and physical aggression and noncompliance. Families' adaptation to this developmental transition forms the basis for subsequent developmental stages (Shaw, Bell, & Gilliom, 2000; Shaw & Gross, in press).

It is increasingly clear that parenting practices are at the core of early-onset problem behavior and adjustment problems and are integral to the solution

---

This research was supported by Grant 5 R01 DA16110 from the National Institutes of Health to the first and second authors. We gratefully acknowledge Cheryl Mikkola for her support in the preparation of this article, the Early Steps and the Child and Family Center staff for their assistance, and all the families who participated in this project.

Correspondence concerning this article should be addressed to Thomas J. Dishion, Child and Family Center, Department of Psychology, University of Oregon, 195 W. 12th Avenue, Eugene, OR 97401. Electronic mail may be sent to dishion@uoregon.edu.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7905-0012

(Dishion & Stormshak, 2007). Negative and neglectful parenting practices when children are age 2 are prognostic of later problem behavior, whereas the predictive validity of the child's behavior emerges somewhat later in development (Shaw & Gross, in press). It is clear that the covariation between early parenting practices, children's problem behavior, and later adjustment problems is multidetermined, including factors such as temperament, social context, and proximal factors describing the parents' adjustment (Shaw et al., 2000). Specifically, we know that among children assessed to be at genetic risk for problem behavior, punitive and harsh parenting practices greatly exacerbate that risk (Caspi et al., 2002). Studies of the genetic contributions to the development of antisocial behavior suggest a vulnerability model, in that some children are simply more vulnerable to pathogenic environments. The upside of this conclusion is that vulnerable children, in theory, will be more responsive to positive changes in the environment, such as those effected by systematic interventions. Intervention research with young adults diagnosed with liability for schizophrenia reveal that an emphasis on positive family management reduces the negative affective tone of the family and the mental health difficulties associated with the disorder (Falloon et al., 1985).

Thus, it is important to go beyond telling parents what they should not do (e.g., use harsh punitive parenting) and to identify and promote positive parenting practices that prevent the development of children's problem behavior and replace negative parenting practices. In their children's early years, parents can support the development of the children's competence by using a combination of parenting practices that includes a warm, trusting relationship (Shaw et al., 2000), attentive involvement (Glik, Greaves, Kronenfeld, & Jackson, 1993), positive reinforcement for skill development (Forgatch & Toobert, 1979; Supplee, Unikel, & Shaw, 2007), and proactively structuring situations to promote the development of self-regulation and minimize problem behavior (Gardner, 1994; Gardner, Sonuga-Barke, & Sayal, 1999). Positive parenting strategies such as these are consistent with what has been referred to as *positive behavior support strategies*, a prevalent and effective educational behavior management principle that emphasizes the use of nonaversive, reinforcing adult–child interactions to promote development (e.g., Horner & Carr, 1997; Sugai, Horner, & Sprague, 1999).

A key early intervention strategy, therefore, is to strengthen parents' use of positive behavior support strategies in early childhood to manage and prevent common toddler problem behaviors (Gardner et al., 1999; Gardner, Ward, Burton, & Wilson, 2003; Sanders, 1999). Consistent with this notion, Pettit and Bates (1989) found increased parent–child play and social contact in the 1st and 2nd years of a child's life to be associated with fewer conduct problems at age 4. Similarly, Gardner et al. (1999) found that if parents of toddlers timed their positive interaction strategies to anticipate troublesome situations, 2 years later the risk of child behavior problems was diminished. Successfully engaging parents in positive behavior support practices may help increase the frequency of seemingly mundane parent–child interactions such as conversation and play, which are formative to language development and self-regulation (Baldwin, 1995; Baldwin, Bill, Desjardins, Irwin, & Markman, 1996; Hart & Risley, 1995).

As shown in Figure 1, the infusion of positive behavior support strategies in early development is thought to reduce the development of problem behavior and thereby decrease the likelihood of reactive parenting strategies and coercive escalations (Patterson, 1982; Snyder, Edwards, McGraw, Kilgore, & Holten, 1993). Parent–child coercive interactions serve to increase the likelihood of new forms of child misbehavior and more serious forms of antisocial behavior (Dumas, Lemay, & Dauwalder, 2001; Patterson, Reid, & Dishion, 1992). These early-onset problem behaviors also tend to disrupt children's development of positive skills and adjustment to school (Campbell, Pierce, Moore, Marakovitz, & Newby, 1996; Dishion, 1990; Ingoldsby, Shaw, & Garcia, 2001; Shaw, Owens, Vondra, Keenan, & Winslow, 1996). Accordingly, problem behavior in middle childhood as identified by teachers or parents is prognostic of continuity to adolescence, including more serious antisocial activity (e.g., arrest), substance use, and

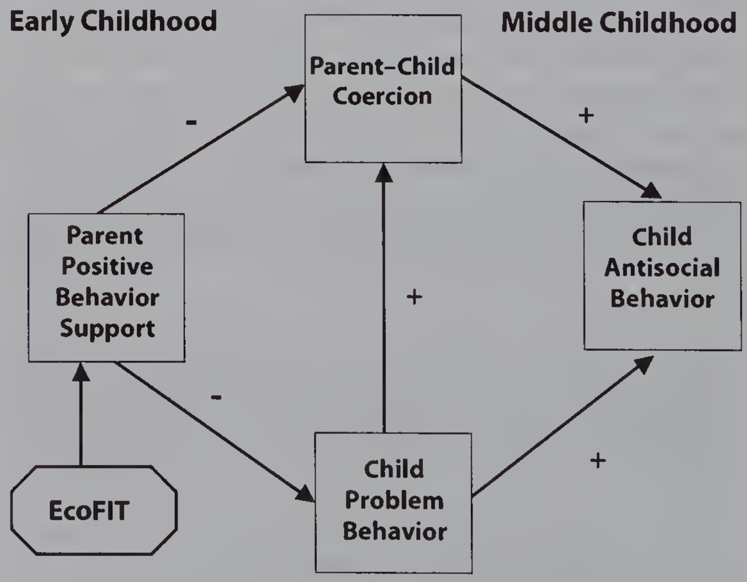


Figure 1. Theoretical framework for the early steps intervention.

sexual activity begun between ages 11 and 15 (Loeber & Dishion, 1983; Patterson et al., 1992; Shaw & Gross, in press). In addition to exacerbating child problem behavior, coercive parent–child interactions and a history of child antisocial behavior often undermine the parent–child relationship (Patterson & Dishion, 1988), making efforts to intervene at later developmental stages more challenging because parents may have "given up" their efforts to influence and socialize their child (Dishion, Nelson, & Bullock, 2004; Dishion & Patterson, 1992).

Family Interventions

If a small subset of children has the highest rate of problem behavior and psychopathology within a community and these risk factors can be reliably identified in early childhood, intervention programs logically would prioritize these children's families. If effective, the prevention efforts would improve parent–child interactions and reduce risk among the most vulnerable youth, on a variety of adjustment indices. The Family Check-Up (FCU) intervention specifically targets disrupted and unskilled family management practices in early childhood to reduce and prevent later problem behavior and focuses on parenting factors and variables found to compromise parenting quality (e.g., temperament, parental well-being) that have also been shown to be directly related to trajectories of persistent problem behavior. The term *family management* describes a collective set of parenting skills known to be highly related to child success (Bullock & Forgatch, 2005; Forgatch, Bullock, & Patterson, 2004; Patterson, 1982; Patterson et al., 1992).

That family-centered strategies are effective for reducing child problem behavior is strongly supported in the treatment and prevention science literatures across developmental periods (e.g., Brody et al., 2004; Conduct Problems Prevention Research Group, 2002; Dishion & Patterson, 1992; Eddy & Chamberlain, 2000; Forgatch & DeGarmo, 1999; Henggeler, Schoenwald, Borduin, Rowland, & Cunningham, 1998; Kazdin, 2003; Liddle, 1999; Sanders, 1999; Spoth, Redmond, & Shin, 1998), including early childhood (Brinkmeyer & Eyberg, 2003; Olds, 2002; Olds et al., 1997; Webster-Stratton, 1990). Family-centered interventions have undergone a critical shift in the past 20 years, moving from a treatment model that is delivered to clients in clinic settings to a prevention model involving proactive recruitment of parents to engage in interventions in home and community settings such as schools, day care, and community support services (Gardner, Burton, & Klimes, 2006; Hutchings et al., 2007; Shaw, Dishion, Supplee, Gardner, & Arnds, 2006).

It is not difficult to argue that the earlier one intervenes with a family, the easier it is to make significant developmental changes and reduce the likelihood of early-starting conduct problems and later drug-use risk. There are three reasons for the early intervention hypothesis. First, early interventions target the child behavior before more serious forms of antisocial behavior develop. Noncompliant and oppositional behavior is easier to remediate than are lying, stealing, and proactive aggression. Targeting behavior early is expedient because a child's ecology expands with development and moves from home into school and neighborhood settings that are less predictable and that involve relationships with teachers (Pianta, Steinberg, & Rollins, 1995) and same-age and older peers (Dishion, Duncan, Eddy, Fagot, & Fetrow, 1994; Ingoldsby & Shaw, 2002). Eventually, peers may transform reactive aggression into proactive aggression and may support other forms of negative behavior (Poulin & Boivin, 2000). By adolescence, peers can become a powerful context within which delinquency, drug use, and sexual experimentation are embedded (Capaldi, Crosby, & Stoolmiller, 1996; Dishion & Owen, 2002; French & Dishion, 2003; Patterson, Dishion, & Yoerger, 2000). Thus, family interventions initiated in early childhood are more likely to meet the objective of preventing a variety of later behavior problems.

The second advantage of early intervention is that parents are probably younger and have undergone fewer development transitions themselves, including having multiple children and multiple partner transitions, and they may likely engage in fewer high-risk behaviors that compromise their own health. In particular, divorce and remarriage have disruptive effects on the parent–child relationship and child adjustment, and although these transitions can be opportune times to intervene (e.g., Martinez & Forgatch, 2001), it certainly would be better to prevent their occurrence in the first place. Early interventions by Olds and colleagues not only reduced the number of caregiver relationship transitions but also the number of ensuing children (Olds, Hill, Robinson, Song, & Little, 2000; Olds et al., 1997).

Third, the sense of optimism caregivers have regarding the possibility of parent–child relationship change is much higher during their offspring's early childhood. As noted earlier, parents' motivation to change may be significantly reduced after years of predominantly acrimonious interchanges with a child compared with when they are parents of toddler-age offspring. We found that parents of adolescents were more likely to drop out of behavioral family therapy treatment than were parents seeking treatment with younger children (Dishion & Patterson, 1992). In the FCU model, only 25% of the parents of adolescents engaged in the intervention (Connell, Dishion, Yasui, & Kavanagh, 2007), whereas more than 75% of families in early childhood trials engaged in the FCU (Shaw et al., 2006).

The empirical evidence for the effectiveness of family-centered interventions during early childhood is impressive. The pioneering work of Webster-Stratton and colleagues is exemplary. By using videotaped examples and promoting the use of systematic, behavioral therapy-based parenting skills that build a productive parent–child relationship and strengthen family management, significant improvements have been observed and replicated (e.g., Webster-Stratton & Taylor, 2001). In the family management model, the interventions are offered in a parent group format, and caregivers' discussions about the videotaped examples are a compelling source of change.

The Ecological Approach to Family Intervention and Treatment Model

Despite these promising findings, current forms of empirically supported family interventions arguably require more efficient and brief interventions that can be rolled out via service delivery systems that have contact with large numbers of children and families (Hoagwood & Koretz, 1996). The ecological approach to family intervention and treatment (EcoFIT) was designed with a public health focus to ensure that the needs of a wider range of children and families are met, from prevention to treatment. The focus also ensures efficiency in terms of cost and in delivery in settings that reach a large number of children, such as public school systems (Dishion & Stormshak, 2007). The intervention is "ecological" in that it is designed to improve children's adjustment across settings (home, school, and neighborhood) by motivating positive behavior support practices in those settings. Two key features of the intervention are that it is assessment driven and tailored to the needs of youth and families as revealed by family observations, assessments in extrafamilial contexts (e.g., schools), and reports from important individuals in the child's life (e.g., parents, teachers, youth when of age), and that typically, intervention is brief.

The EcoFIT model is predicated on the FCU (Dishion & Kavanagh, 2003), which is a brief parenting intervention that emphasizes motivation to change (Miller & Rollnick, 2002). The FCU involves three sessions. The first step is a meeting with parents to explore their perceptions and concerns regarding their family setting and child's behavior. The second step is a comprehensive assessment that includes videotaping parent–child interactions. The third step is a structured feedback session that is based on the results of the assessment and that emphasizes parenting and family strengths yet draws attention to possible areas of change. The intervention is motivational in that it stimulates caregivers to address key problems in parenting either on their own or with the support of a professional. Interventions that follow the FCU are thus tailored to each family's needs on the basis of the assessment and of the parents' motivation to change. Consequently, some parents may focus on only one dimension of their parenting practices, unlike with conventional parenting intervention programs that tend to emphasize a standard curriculum of parenting practices to all parents.

The tailored, adaptive, and motivational aspect of the intervention results in relatively brief interventions. The application of the EcoFIT to families of young adolescents within the school setting revealed that an average of six sessions for the highest risk families during the course of 2–3 years resulted in enduring reductions in drug use and problem behavior and improved parental monitoring practices (Connell et al., 2007; Dishion, Nelson, & Kavanagh, 2003).

In pilot work leading to this study, we applied the FCU intervention to high-risk families of toddlers involved in a national family nutrition and health program for young families referred to as the Women, Infants, and Children Nutrition Program (WIC). We randomly assigned 120 families of toddlers to either WIC as usual or WIC with one FCU at child age 2 and found that the intervention reduced subsequent problem behavior and improved parent involvement at child ages 3 and 4, respectively (Shaw et al., 2006). Moreover, videotaped home observation sessions showed that random assignment to the FCU had resulted in improvements in caregivers' use of proactive parenting practices and reduced child negative behavior at age 3 (Gardner, Shaw, Dishion, Burton, & Supplee, 2007). These findings generally fit with those of two dissertation projects that used the FCU with families with toddlers and revealed improved parenting and reduced child problem behavior (Jones, 2004; O'Leary, 2000).

The WIC study of the FCU was limited in terms of sample size, intervention services offered to the families, and use of only male children from an urban community. Our current study, referred to as the Early Steps Multisite Study, remedies these limitations. First, the sample includes 731 at-risk families; half were randomly assigned to the FCU/EcoFIT intervention and the remainder to WIC as usual. Second, the families were recruited from three geographically and culturally diverse regions, including metropolitan Pittsburgh, PA; suburban Eugene, OR; and rural Charlottesville, VA. The sample also reflects cultural diversity, including African American, European American, and Latino families. Third, we used the entire EcoFIT model in that families were provided additional services following the FCU, consistent with an adaptive, tailored approach to intervention. We examined outcomes on parent reports of child problem behavior and direct observations of positive behavior support practices, including positive reinforcement, proactive parenting, parent involvement, and joint-attention verbal interaction.

In this study, we hypothesize that families at high risk involved in WIC and randomized to the EcoFIT would (a) report reductions in problem behavior from child age 2 to 4 years compared with control families, (b) show increases in caregiver involvement and direct observations of positive behavior support practices at child ages 2 and 3 compared with control participants, and (c) show that reductions in problem behavior would be mediated by improvements in positive behavior support practices, as measured by home visitor ratings and direct observations of parent–child interaction.

Method

Participants

Participants included 731 mother–child dyads recruited between 2002 and 2003 from WIC programs in the metropolitan areas of Pittsburgh, PA, and Eugene, OR, and in and outside the town of Charlottesville, VA. Families were contacted at WIC sites and invited to participate if they had a son or daughter between age 2 years 0 month and 2 years 11 months, following a screen to ensure that they met the study criteria by having socioeconomic, family, and/or child risk factors for future behavior problems. Risk criteria for recruitment were defined at or above 1 *SD* above normative averages on several screening measures in the following three domains: (a) child behavior (conduct problems, high-conflict relationships with adults), (b) family problems (maternal depression, daily parenting challenges, substance-use problems, teen parent status), and (c) sociodemographic risk (low education achievement and low family income, relevant to WIC criterion). Two or more of the three risk factors were required for inclusion in the sample.

Recruitment. As shown in Figure 2 and in Table 1, of the 1,666 families who had children in the appropriate age range and who were contacted at WIC sites across the three study sites, 879 met the eligibility requirements (52% in Pittsburgh, 57% in Eugene, and 49% in Charlottesville) and 731 (83.2%) agreed to participate (88% in Pittsburgh, 84% in Eugene, and 76% in Charlottesville). The children in the sample had a mean age of 29.9 months (*SD* = 3.2) at the time of the age 2 assessment.

Of the 731 families (49% female), 272 (37%) were recruited in Pittsburgh, 271 (37%) in Eugene, and 188 (26%) in Charlottesville. Across sites, primary caregivers self-identified as belonging to the following ethnic groups: 28% African American, 50% European American, 13% biracial, and 9% other groups (e.g., American Indian, Native Hawaiian). Thirteen percent of the sample reported being Hispanic American. During the 2002–2003 screening period, more than two thirds of those families enrolled in the project had an annual income of less than $20,000, and the average number of family members per household was 4.5 (*SD* = 1.63). Forty-one percent of the population had a high school diploma or general education diploma (GED), and an additional 32% had 1–2 years of post–high school training.

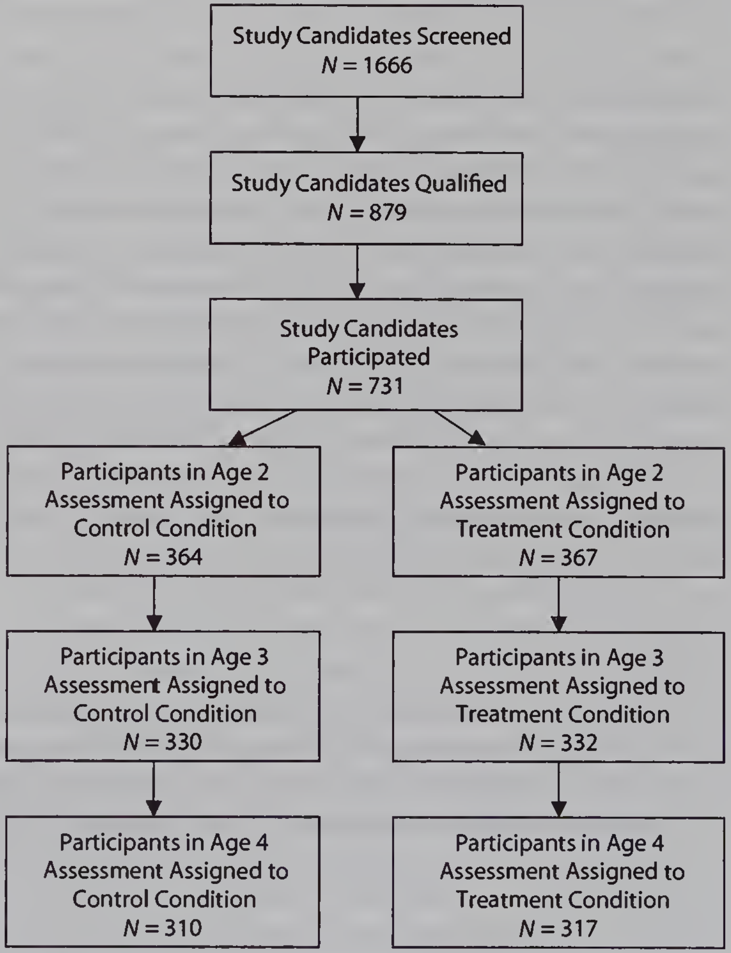

Study Candidates Screened
N = 1666
Study Candidates Qualified
N = 879
Study Candidates Participated
N = 731
Participants in Age 2 Assessment Assigned to Control Condition
N = 364
Participants in Age 2 Assessment Assigned to Treatment Condition
N = 367
Participants in Age 3 Assessment Assigned to Control Condition
N = 330
Participants in Age 3 Assessment Assigned to Treatment Condition
N = 332
Participants in Age 4 Assessment Assigned to Control Condition
N = 310
Participants in Age 4 Assessment Assigned to Treatment Condition
N = 317

Figure 2. An overview of the recruitment, assessment, and randomization protocol.

Table 1
Recruitment Descriptives by Project Site

	Project site			
	Pittsburgh	Eugene	Charlottesville	Total sample
Recruitment (*n*)				
Screened	596	565	505	1,666
Qualified	309	323	247	879
Participated	272	271	188	731
Participant demographics (%)				
Race				
African American	50.4	1.5	33.5	27.9
European American	38.1	70.0	39.4	50.1
Biracial	10.0	23.5	15.4	13.0
Other race	1.5	5.0	11.7	8.9
Ethnicity				
Hispanic	1.8	20.0	20.7	13.4
Target child age, *M* (*SD*)	28.3 (3.49)	28.5 (2.91)	27.7 (3.43)	28.2 (3.28)
Target child gender	49.6% female	49.8% female	48.9% female	49.5% female
Annual family income < $20,000 (%)	70.5	62.4	66.0	66.3
Family members per household, *M* (*SD*)	4.4 (1.55)	4.5 (1.67)	4.6 (1.66)	4.5 (1.63)
Education (%)				
High school diploma	42.5	39.5	40.0	41.0
1–2 years post–high school	35.7	34.7	25.5	32.7
Treatment participation (%)				
Age 2 feedback received	76.5	78.7	78.9	77.9
Age 3 feedback received	66.6	70.4	56.3	65.4
Age 4 feedback received	66.6	71.9	53.2	65.3

Retention. Of the 731 families who initially participated, 659 (90%) were available at the 1-year follow-up and 619 (85%) participated at the 2-year follow-up when children were between 4 years and 4 years 11 months old. At ages 3 and 4, selective attrition analyses revealed no significant differences in project site; children's race, ethnicity, or gender; levels of maternal depression; or children's externalizing behaviors (parent reports). Furthermore, no differences were found in the number of participants who were not retained in the control versus the intervention groups at both age 3 ($n = 40$ and $n = 32$, respectively) and age 4 ($n = 58$ and $n = 53$, respectively).

Measures

Demographics questionnaire. A demographics questionnaire was administered to the mothers during the child ages 2, 3, and 4 visits. This measure included questions about family structure, parental education and income, parental criminal history, and areas of familial stress.

Maternal depression. We used a brief but valid indicator of adult depression referred to as the Center for Epidemiological Studies on Depression Scale (CES–D; Radloff, 1977). The CES–D is a well-established and widely used 20-item measure of depressive symptomatology that was administered to mothers at the child ages 2 and 3 home assessments. Participants report how frequently they have experienced the listed depressive symptoms during the past week on a scale ranging from 0 (*less than 1 day*) to 3 (*5–7 days*). Items are summed to create an overall depressive symptoms score. For the current sample, internal consistencies were .76 and .75 at the ages 2 and 3 assessments, respectively.

Early childhood problem behavior. The Child Behavior Checklist (CBCL) for ages 1.5–5 (Achenbach & Rescorla, 2001) is a 99-item questionnaire that assesses behavioral problems in young children. Mothers completed the CBCL at the child ages 2, 3, and 4 home assessments. The CBCL includes one broadband factor that assesses externalizing symptoms, *Externalizing*, which was used as the primary outcome measure in our study. Internal consistencies for Externalizing were .86, .89, and .86 at ages 2, 3, and 4, respectively, in the current study.

At child ages 2, 3, and 4 assessments, we administered the Eyberg Child Behavior Inventory, a widely used 36-item measure of early childhood problem behavior (Robinson, Eyberg, & Ross, 1980). The Eyberg includes two factors that focus on the perceived intensity of a behavior and degree to which the behavior is a problem for caregivers. Because the Intensity factor is similar in content and structure to the CBCL Externalizing factor, for our study, we focused on the Problem factor, which asks caregivers to use a 7-point scale to report the extent to which the behavior is a problem for the parent. The inventory has been demonstrated to be highly correlated with independent observations of children's behavior, to differentiate clinic-referred and nonclinic populations (Robinson et al., 1980) and to show high test–retest reliability (.86) and internal consistency (.98; Webster-Stratton, 1985). In our study, internal consistencies for the Problem factor were .84, .90, and .94 at ages 2, 3, and 4, respectively.

Procedures

Assessment protocol. Parents (i.e., mothers and, if available, alternative caregivers such as fathers or grandmothers) and children who agreed to participate in the study were scheduled for a 2.5-hr home visit. Each assessment began by introducing children to an assortment of age-appropriate toys and having them play for 15 min while the mothers completed questionnaires. After the free play (15 min), which began with the child being approached by an adult stranger (i.e., undergraduate videographer), each primary caregiver and child participated in a cleanup task (5 min), followed by a delay of gratification task (5 min), four teaching tasks (3 min each, with the last task being completed by alternate caregiver and child), a second free play (4 min), a second cleanup task (4 min), the presentation of two inhibition-inducing toys (2 min each), and a meal preparation and lunch task (20 min). Exactly the same home-visit and observation protocol was repeated at ages 3 and 4 for both the control and the intervention groups. Families received $100 for participating in the age 2 home visit. Families were reimbursed $120 at the age 3 assessment and $140 at the age 4 assessment.

The randomization sequence was computer generated by a staff member who was not involved with recruitment. Randomization was gender balanced to ensure an equal number of males and females in the control and intervention subsamples. To ensure that the randomization was blinded, the examiner opened a sealed envelope, revealing the family's group assignment only after the assessment was completed, and shared this information with the family. Examiners carrying out follow-up assessments were not informed of families' randomly assigned condition.

Relevant to this study, we present maternal reports of child externalizing problems from ages 2, 3, and 4 assessments and direct observations of parent–child interaction from the child ages 2 and 3 assessments. Coding of family assessment videotapes was completed for ages 2 and 3 to test the hypothesis of a link between improvements in positive behavior support in response to the intervention and change in young children's problem behavior.

Coding of videotaped parent–child interactions. A team of undergraduates coded the videotaped family interaction tasks by using the Relationship Process Code (RPC; Jabson, Dishion, Gardner, & Burton, 2004). The average team RPC percent agreement = .87, κ = .86. The RPC is a third-generation code derived from the Family Process Code (Dishion, Gardner, Patterson, Reid, & Thibodeaux, 1983) used extensively in previous research. After coding each family interaction, coders completed a coder impressions inventory regarding proactive and positive behavior support practices, for the purpose of this research study. All family interaction tasks were evaluated in the scoring of positive behavior support practices. In addition, the home visitors' ratings of parent involvement with the young child were used as another indicator of the positive behavior support construct. Although coders were predominantly European American (90%), protocols developed by using examples of culturally diverse coding categories and by extensive training ensured that coding of family interactions was culturally sensitive. Our previous research revealed that cultural biases in coding of African American family interactions existed when coders were untrained in the coding system and that coder training resulted in eliminations of coding differences between European American and African American coders (Yasui & Dishion, 2007).

In detail, the following items were entered into the positive behavior support scores:

1. Parent Involvement. This measure is based on the home visitor's rating of the parents' involvement, which used the following items from the Home Observation for Measurement of the Environment inventory (HOME; Bradley, Corwyn, McAdoo, & Garcia-Coll, 2001): "Parent keeps child in visual range, looks at often"; "Parent talks to child while doing household work"; and "Parent structures child's play periods."
2. Positive Behavior Support. This measure is based on videotape coding (durations) of caregivers prompting and reinforcing young children's positive behavior as captured in the

following RPC codes: positive reinforcement (verbal and physical), prompts and suggestions of positive activities, and positive structure (e.g., providing choices in a request for behavior change).

3. Engaged Parent–Child Interaction Time. This measure reflects the average length of parent–child sequences that involve talking or physical interactions such as turn taking or playing a game. Thus, the average duration of episodes that included consecutive parent–child exchanges involving RPC codes such as Talk and Neutral Physical Contact were used to define these episodes.
4. Proactive Parenting. Videotape coders rated each parent on his or her tendency to anticipate potential problems and to provide prompts or other structural changes to avoid young children becoming upset and/or involved in problem behavior on the following six items: parent gives child choices for behavior change whenever possible; parent communicates to the child in calm, simple, and clear terms; parent gives understandable, age-appropriate reasons for behavior change; parent adjusts/defines the situation to ensure the child's interest, success, and comfort; parent redirects the child to more appropriate behavior if the child is off task or misbehaves; parent uses verbal structuring to make the task manageable ($\alpha = .835$).

Intervention protocol: The FCU. Families randomly assigned to the intervention condition were then scheduled to meet with a parent consultant for two or more sessions, depending on the family's preference. The FCU is a brief, three-session intervention based on motivational interviewing and modeled after the Drinker's Check-Up (Miller & Rollnick, 2002). Typically, the three meetings include an initial contact session, an assessment session, and a feedback session (Dishion & Kavanagh, 2003). However, to optimize the internal validity of the study (i.e., prevent differential drop out for experimental and control conditions), the assessments were completed before random assignment results were known to either the research staff or the family. Thus, for the purpose of research studies only, the sequence of contacts comprised an assessment (baseline), randomization, an initial interview, a feedback session, and possibly follow-up sessions. Families were given a gift certificate for $25 for completing the FCU at the end of the feedback session, which could be used at local supermarkets or video stores.

As described earlier, the initial meeting was an assessment conducted with research staff, during which the family engaged in a variety of in-home videotaped tasks of parent–child interaction and caregivers completed several questionnaires about their own, their child's, and their family's functioning. Also during this home assessment, staff completed ratings of parent involvement with and supervision of their child. The second session was an initial interview by the parent consultant, during which the consultant explored parent concerns, focusing on family issues that were currently the most critical to the child's well-being. The third meeting involved a feedback session, during which the parent consultant summarized the results of the assessment by using motivational interviewing strategies. An essential objective of the feedback session is to explore the parents' willingness to change problematic parenting practices, to support existing parenting strengths, and to identify services appropriate to the family needs. The parents were also offered follow-up sessions that focused on parenting practices, other family management concerns (e.g., coparenting), and contextual issues (e.g., child care resources, marital adjustment, housing, vocational training).

Parent consultants who completed the FCU and follow-up parenting sessions were a combination of Ph.D.- and master's-level service workers, all of whom had previous experience in carrying out family-based interventions but at the study's outset had no experience in using the FCU. The parent consultants were of diverse ethnic status, including Latino, African American, European American (the majority), and mixed ethnicity. The consultants were initially trained for 2.5–3 months by using a combination of strategies, including didactic instruction and role playing, followed by ongoing videotaped supervision of intervention activity. Before working with study families, parent consultants were initially certified by lead parent consultants at each site who had been certified by the intervention developer (Dishion, 1990). Certification was established by reviewing videotapes of feedback and follow-up intervention sessions to evaluate whether the parent consultants were competent in all critical components of the intervention as described later in this article. This process is repeated yearly to reduce drift from the intervention model and adheres to the methods of Forgatch, Patterson, and DeGarmo (2005), which revealed that direct observations of therapist fidelity to parent management training predicted change in parenting practices and child behavior. In addition, cross-site case reviews were convened weekly via videoconference to further enhance fidelity. Finally, annual parent consultant meetings were held to update training, discuss possible changes in the intervention model, and address special intervention issues reflected by the needs of families across sites.

Of the families assigned to the treatment condition, 77.9% participated in the initial interview and feedback sessions at child age 2, and 65.4% participated at child age 3 (see Table 1 for site-specific data). Of those families who met with a parent consultant, the average number of sessions per family was 3.32 (*SD* = 2.84). The number of sessions was uncorrelated with future levels of problem behavior after controlling for initial levels.

Data analysis. We used a variety of strategies to test the hypotheses guiding this research. Latent growth mixture modeling (LGMM) analyses were conducted using Mplus 4.21 (Muthén & Muthén, 2007) to examine heterogeneity in developmental trajectories of early problem behaviors and positive parenting, including possible heterogeneity in intervention response. Three separate sets of analyses were conducted, examining possible heterogeneity in (a) Eyberg Problem Behavior trajectories, (b) CBCL Externalizing behavior trajectories, and (c) change in the Positive Parenting latent variable. Mixture modeling is an active area of methods research regarding the optimal approach to determining the number of latent classes (e.g., Nylund, Asparouhov, & Muthén, 2006). We attempted to obtain fit indices for conditional models with one to six classes at each age. Models with different numbers of latent classes are not nested; therefore, Muthén and Muthén (2000) have recommended the following four criteria for selecting the optimal number of latent classes in factor mixture models: (a) the Bayesian information criterion (BIC) and a sample-size-adjusted version of the BIC (adj. BIC), with lower scores representing better fitting models; (b) the quality of classification across models, represented by entropy, with higher entropy values indicating better classification of individuals into their most likely trajectory class; (c) the Lo–Mendel–Rubin–likelihood ratio test (LMR–LRT) and the bootstrap likelihood ratio test (BLRT), each of which provides a statistical comparison of the fit of a given model with a model of one fewer classes; and (d) the theoretical relevance and usefulness of latent trajectory classes. Recent simulation studies by Nylund et al. (2006) supported the use of adj. BIC and BLRT for selection of the optimal number of classes in latent class analysis models, with the BLRT providing particularly consistent correct results. In light of these findings, primary weight was placed on the BLRT and the adj. BIC values in selecting the number of classes. All analyses were conducted using 100 randomized start values run for 10 iterations each, with the best fitting 25 randomized start values run to model convergence.

Subsequent mediator analyses examined the indirect effect of intervention on the rate of change in problem behaviors through the effect of intervention on maternal positive behavior support at child age 3. The basic mediation model is shown in Figure 3. In the mediation models, the slope of problem behaviors was regressed on age 3 positive behavior support and intervention status, and positive behavior support at child age 3 was regressed on intervention status, the initial level of child problem behaviors, and positive behavior support at child age 2. Thus, this model tests whether intervention is related to the change in positive behavior support from child ages 2 to 3, and whether this change in positive behavior support, in turn, predicts the rate of change in child behavior problems from ages 2 to 4, controlling for the direct effect of intervention. A statistical test of the significance of the indirect effect from intervention to the change in maternal symptoms to the rate of change in problem behavior was examined, with standard errors for indirect effects calculated using the delta method described by MacKinnon and colleagues (MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002; MacKinnon, Lockwood, & Williams, 2004).

Results

Descriptive statistics for all variables are shown in Table 2. For ease of interpretation, we present *t* scores on the Eyberg and CBCL measures, although raw scores were used for models to avoid potential age and gender corrections. The percentage of the respondents in the clinical range on these measures at each age is also presented in Table 2. In terms of validating children's problem behavior status, for both the CBCL Externalizing and the Eyberg Problem Behavior factors, mean scores were approximately 1 *SD* above normative scores at age 2. Using the borderline clinical cutoff of the 90th percentile for the CBCL, 48.6% of children were reported to have clinically elevated scores on the Externalizing factor at age 2. This percentage was reduced over time to 24% at age 4.

Correlations for problem behavior variables from age 2 to 4 and correlations for positive behavior support variables from age 2 to 3 are shown in Table 3. It is important to note that no significant associations were found between treatment group and child gender or ethnicity, levels of maternal depressive symptoms, or any type of child problem

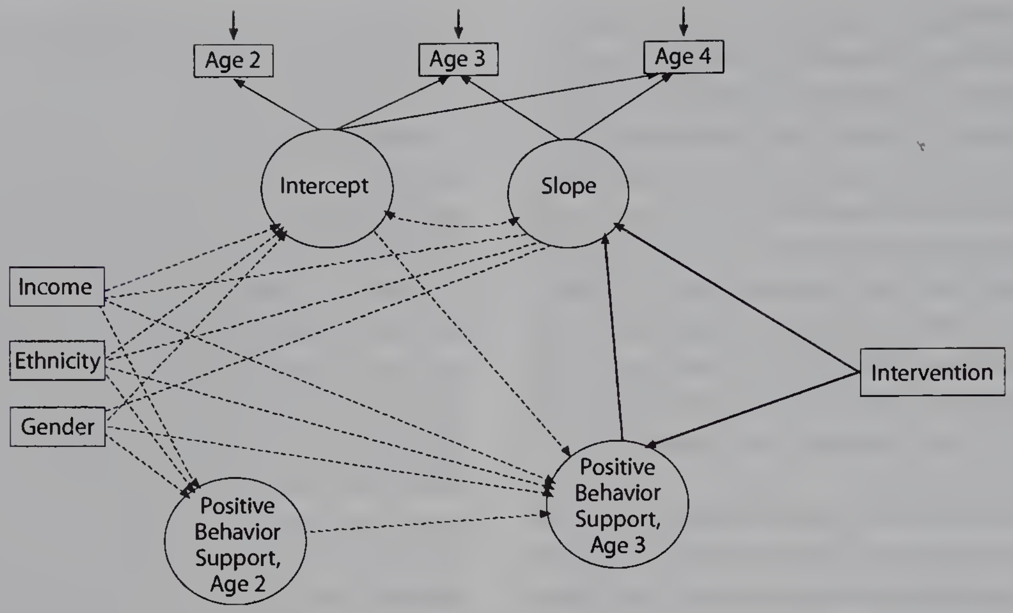

Age 2
Age 3
Age 4
Intercept
Slope
Income
Ethnicity
Gender
Intervention
Positive Behavior Support, Age 3
Positive Behavior Support, Age 2

Figure 3. Mediation model overview: Test of indirect path from treatment to age 3 positive behavior support to slope of problem behavior.

behavior at age 2, suggesting that randomization was successful. Modest to moderate associations were consistently found among the observed indicators of positive behavior support. In addition, modest but significant negative associations were found between ethnicity and dimensions of positive behavior support, with African American mothers observed to show lower levels than European American mothers.

The central analyses tested the following hypotheses: (a) that random assignment to a family-centered intervention would result in reductions in the frequency of externalizing problems and those problems being perceived as problematic by mothers, (b) that intervention would also result in improvement in parenting practices, and (c) that reduction in the frequency of externalizing symptoms and those symptoms that were a problem for mothers would be mediated by improvements in parenting. All analyses were conducted using Mplus 4.1 and used full information maximum likelihood estimation (Muthén & Muthén, 2004), which provides a method for accommodating missing data by estimating each parameter by using all available data for the estimation of that specific parameter.

Table 2
Descriptive Statistics

	M	*SD*	Total *N*	*N* (%) in clinical range
Eyberg Problem Behavior (*t* score), age 2	59.18	8.46	729	323 (44.2)
Eyberg Problem Behavior (*t* score), age 3	59.61	10.36	642	321 (43.9)
Eyberg Problem Behavior (*t* score), age 4	59.63	11.01	616	302 (41.3)
CBCL Externalizing behavior (*t* score), age 2	59.49	8.21	730	355 (48.6)
CBCL Externalizing behavior (*t* score), age 3	55.97	9.39	651	224 (30.6)
CBCL Externalizing behavior (*t* score), age 4	53.66	10.47	619	175 (23.9)
CBCL Internalizing behavior (*t* score), age 2	56.32	8.53	730	282 (38.6)
CBCL Internalizing behavior (*t* score), age 3	54.30	9.61	651	203 (27.8)
CBCL Internalizing behavior (*t* score), age 4	53.25	10.09	619	170 (23.3)

Note. CBCL = Child Behavior Checklist.

Hypothesis 1: Problem Behavior in Early Childhood

For the Eyberg Problem Behavior scale, model fit indices converged to indicate that a two-class LGMM provided optimal fit to the data. Most centrally, both the LMR–LRT and the BLRT were significant for the two-class versus one-class model (LMR–LRT = 51.24, $\Delta df = 11$, $p < .05$; BLRT = 51.24, $\Delta df = 11$, $p < .05$), and the adj. BIC value was lower for the two-class (adj. BIC = 13232.43) versus the one-class model (adj. BIC = 13246.06). Models with three or more classes yielded higher adj. BIC values (three-class adj. BIC = 13239.57) and nonsignificant LMR–LRT and BLRT values (e.g., three-class LMR–LRT = 25.69, $\Delta df = 10$, $p = .20$; BLRT = 25.69, $\Delta df = 10$, $p = .20$).

Entropy for the two-class model was relatively low (.52), indicating that the two classes were not well discriminated. The two classes appeared to be ordered along a severity dimension, with 56% of youth in an elevated-problem class and 44% of youth

Table 3

Correlations for Problem Behavior Variables

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1. Treatment group	—																	
2. Child gender	.00	—																
3. Child ethnicity	.00	−.05	—															
4. Monthly family income	.00	−.02	−.16*	—														
5. Eyberg Problem Behavior age 2	.00	.00	.03	−.01	—													
6. Eyberg Problem Behavior age 3	−.04	.07	−.04	−.09*	.43*	—												
7. Eyberg Problem Behavior age 4	−.09*	.05	−.04	−.14*	.35*	.65*	—											
8. CBCL Externalizing age 2	.02	.06	−.03	−.11*	.38*	.44*	.38*	—										
9. CBCL Externalizing age 3	−.02	.08*	−.04	−.14*	.29*	.67*	.53*	.60*	—									
10. CBCL Externalizing age 4	−.06	.07	−.07	−.02	.24*	.51*	.67*	.51*	.70*	—								
11. HOME Parent Involvement age 2	−.05	−.06	−.13*	.13*	−.08*	−.03	−.06	−.03	−.04	−.03	—							
12. RPC Positive Parenting age 2	.00	.03	−.06	.04	.05	.02	−.08	−.04	−.05	−.14*	.11*	—						
13. RPC Parent Neutral age 2	−.06	.01	−.19*	.08*	−.02	−.02	−.05	−.08	−.07	−.06	.30*	.31*	—					
14. COIMP Proactive Parenting age 2	.02	−.03	−.17*	.08	−.06	−.10*	−.09	−.19*	−.17*	−.15*	.26*	.28*	.39*	—				
15. HOME Parent Involvement age 3	.02	.00	−.13*	.06	−.09*	−.05	−.10*	−.07	−.12*	−.11*	.22*	.17*	.26*	.29*	—			
16. RPC Positive Parenting age 2	.09*	−.06	−.15*	.09*	.03	−.09*	−.16*	−.05	−.12*	−.11*	.12*	.24*	.24*	.22*	.25*	—		
17. RPC Parent Neutral age 2	.08*	.04	−.16*	.04	.05	−.03	−.03	−.04	−.06	−.07	.24*	.26*	.46*	.24*	.34*	.24*	—	
18. COIMP Proactive Parenting age 2	14*	.01	−.20*	.17*	−.01	−.11*	−.13*	−.10*	−.17*	−.16*	.22*	.24*	.30*	.49*	.32*	.27*	.33*	—

Note. CBCL = Child Behavior Checklist; HOME = Home Observation for Measurement of the Environment Inventory; RPC = Relationship Process code; COIMP = Coder Impressions.
$^*p < .05$.

in a lower problem class, with the average class trajectories nearly parallel from ages 2 to 4. Membership in the elevated-problem class was predicted by male gender (logit = −.69, *SE* = .29, $p < .05$). In the elevated class, ethnic minority status predicted both the initial level (β = 3.31, *SE* = 1.27, $p < .05$) and the slope of problem behaviors (β = 3.56, *SE* = 1.30, $p < .05$). Intervention status significantly predicted reduced growth in problem behaviors in the elevated-problems class (β = −1.03, *SE* = .44, $p < .05$; B = −.16). In the lower problems class, ethnic minority status predicted reductions in both the initial level (β = −3.43, *SE* = 1.49, $p < .05$) and the slope of problem behaviors (β = −5.67, *SE* = 1.07, $p < .05$). Intervention status was not significantly related to the problem behavior slope, however (estimate = −.27, *SE* = .38, *ns*). Results for intervention effects in these two classes are depicted graphically in Figure 4.

For the CBCL Problem Behavior scale, model fit indices did not support the existence of heterogeneous trajectory classes, indicating that a one-class LGMM provided the best fit to the data. Most centrally, LMR–LRT and BLRT results indicated that a two-class model did not result in significantly improved model fit over a one-class model (LMR–LRT and BLRT = 55.15, *ns*). Thus, we used a one-class model to test the hypothesis that the intervention reduced child behavior problems on the CBCL.

The latent growth model for the externalizing scale from the CBCL provided excellent fit to the data, $\chi^2(df = 3) = 6.01$, $p = .11$; comparative fit index (CFI) = .99; root mean square error of approximation (RMSEA) = .04; standardized root mean residual (SRMR) = .02. The model yielded significant intercept (estimate = 20.63, *SE* = .27) and slope values (estimate = −2.00, *SE* = .22), as well as significant residual variance in the intercept (estimate = 41.57, *SE* = 3.82) and slope parameters (estimate = 11.24, *SE* = 1.79). The effect of intervention on the rate of change in problem behavior was significant (estimate = −.82, *SE* = .29; β = −.12), with more growth in problem behavior in the control group.

An additional series of latent growth models examined possible gender and ethnicity differences in the effect of intervention on the rate of change in problem behavior. First, to examine possible gender differences in the effect of intervention, we examined the fit of a two-group model comparing results for males and females. The difference in the chi-square value across a model with all paths constrained to be equal across genders, and a model allowing the intervention effect to vary by gender was found to be nonsignificant for both the Eyberg Problem Behavior ($\Delta\chi^2$ = 1.60, Δdf = 1, *ns*) and the Externalizing Behavior

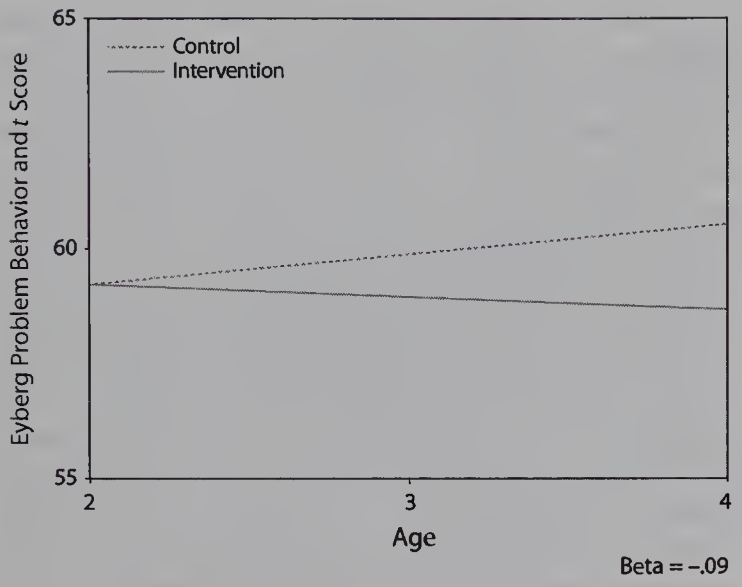


Figure 4. Change from ages 2 through 4 years on the Eyberg measure of child problem behavior (ITT).

Problem scales ($\Delta\chi^2 = 2.21$, $\Delta df = 1$, *ns*), indicating that the intervention effect did not vary by gender.

Second, possible ethnicity differences in the effect of intervention were examined by comparing the fit of models constraining all paths to be equal for European American and ethnic minority families with models allowing the intervention effect to vary across ethnicity groups. The difference in the chi-square value across a model with all paths constrained to be equal across ethnicity groups and a model allowing the intervention effect to vary by ethnicity was found to be nonsignificant for both the Eyberg Problem Behavior ($\Delta\chi^2 = .85$ $\Delta df = 1$, *ns*) and the Externalizing Behavior Problem scales ($\Delta\chi^2 = 2.52$, $\Delta df = 1$, *ns*), indicating that the intervention effect did not vary by ethnicity. In light of the nonsignificant differences in the effects of intervention across either gender or ethnic groups, no further differences in effects across groups were explored.

Hypothesis 2: Observations of Positive Behavior Support Practices

To test Hypothesis 2, we examined the measurement model for the positive behavior support construct, which included the following indicators: (a) home visitor ratings of parent involvement, (b) direct observations of positive behavior support (positive reinforcement, positive prompting, and structuring), (c) direct observation of the engaged parent–child interaction time (parent–child interact time), and (d) coder impressions of proactive parenting.

For the positive parenting latent variable, model fit indices did not support the existence of heterogeneous classes, indicating that a one-class LGMM provided the best fit to the data. Most centrally, LMR–LRT and BLRT results indicated that a two-class model did not result in significantly improved model fit over a one-class model (LMR–LRT and BLRT = 81.20, p = *ns*). Thus, we approached the analysis of the impact of the intervention on positive behavior support by using a one-class, autoregressive model, as shown in Figure 3.

We examined simultaneously the convergent validity of these indicators of parenting and the stability

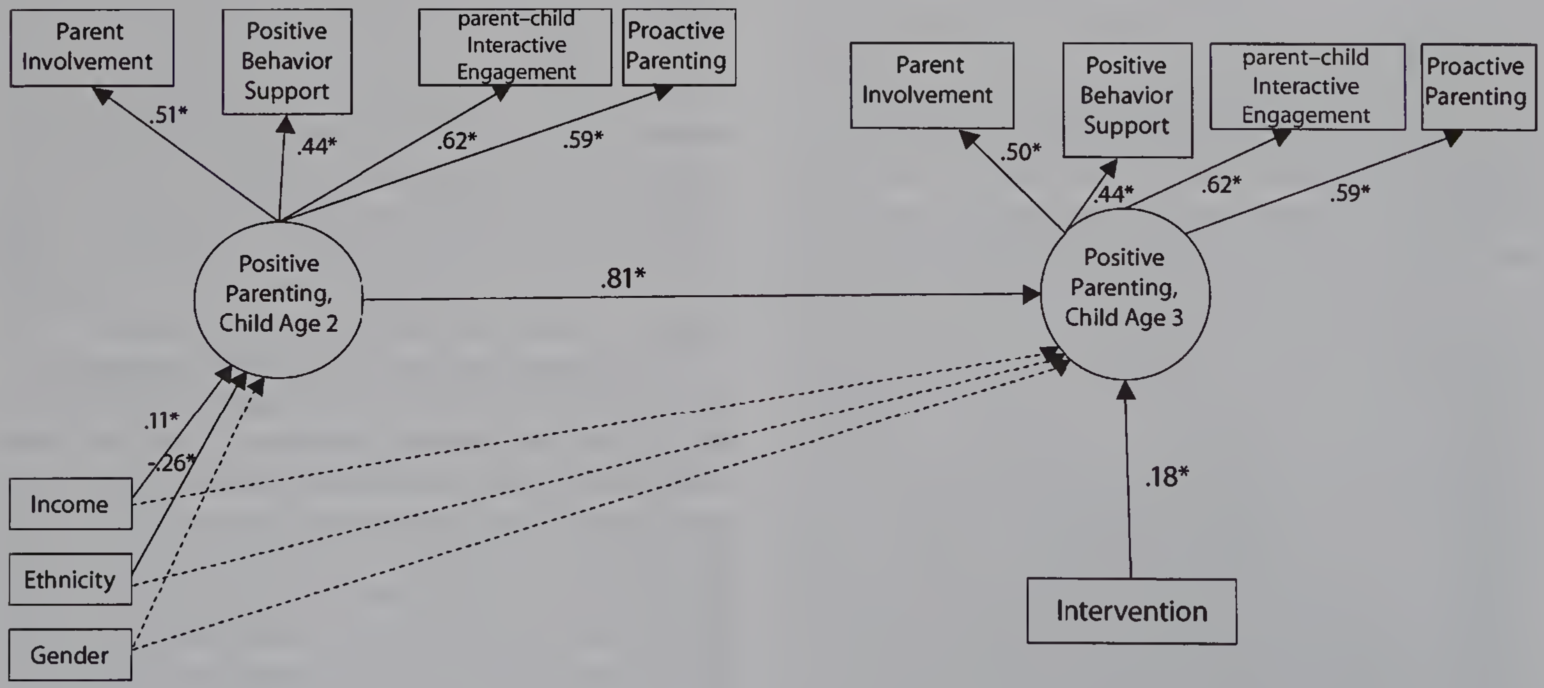

Parent Involvement
Positive Behavior Support
parent–child Interactive Engagement
Proactive Parenting
.51*
.44*
.62*
.59*
Positive Parenting, Child Age 2
.81*
Parent Involvement
Positive Behavior Support
parent–child Interactive Engagement
Proactive Parenting
.50*
.44*
.62*
.59*
Positive Parenting, Child Age 3
.11*
-.26*
Income
Ethnicity
Gender
.18*
Intervention

Figure 5. Positive behavior support: Stability and intervention effect.
Note. P–C = parent-child.
$^*p < .05$.

and the intervention effect, as shown in Figure 5. Factor loadings were constrained to be equal across ages 2 and 3. Figure 5 shows that the four indicators formed a coherent construct and were highly stable from ages 2 to 3 ($\beta = .81$). Despite the high stability, random assignment to the intervention resulted in statistically reliable improvements in observed positive behavior support. The model shown in Figure 3 provides a good fit to the data, $\chi^2(df = 26) = 32.48$, $p = .18$; CFI = .99; RMSEA = .02; SRMR = .04.

Hypothesis 3: Positive Behavior Support and Child Problem Behavior

As indicated, Hypothesis 3 was tested with mediator analyses, examining the indirect effect of intervention on the rate of change in problem behaviors through the effect of intervention on maternal parenting at child age 3. For ease of interpretation, these results are shown in Figures 6 and 7. Because LGMM analyses indicated that there was no heterogeneity in positive behavior support or in CBCL externalizing problems, we tested mediation by using one-class models (i.e., a typical latent variable model including an autoregressive model for positive behavior support and a latent growth model for child behavior problems). As shown in Figure 6, the model for externalizing problem behavior from the CBCL provided reasonable fit to the data by most indices of model fit, $\chi^2(df = 66) = 88.85$, $p = .04$; CFI = .98; RMSEA = .02; SRMR = .04, and the nonsignificant chi-square may be related to the large sample size. In this model, the direct effect of intervention on the problem behavior slope was not significant with maternal parenting included in the model. Treatment significantly predicted improvements in positive behavior support from child ages 2 to 3. More positive behavior support predicted less growth in problem behavior. The indirect effect from intervention to more positive behavior support to decreased growth in problem behavior was statistically significant, although small in magnitude, indicating a significant partial mediation effect of positive behavior support.

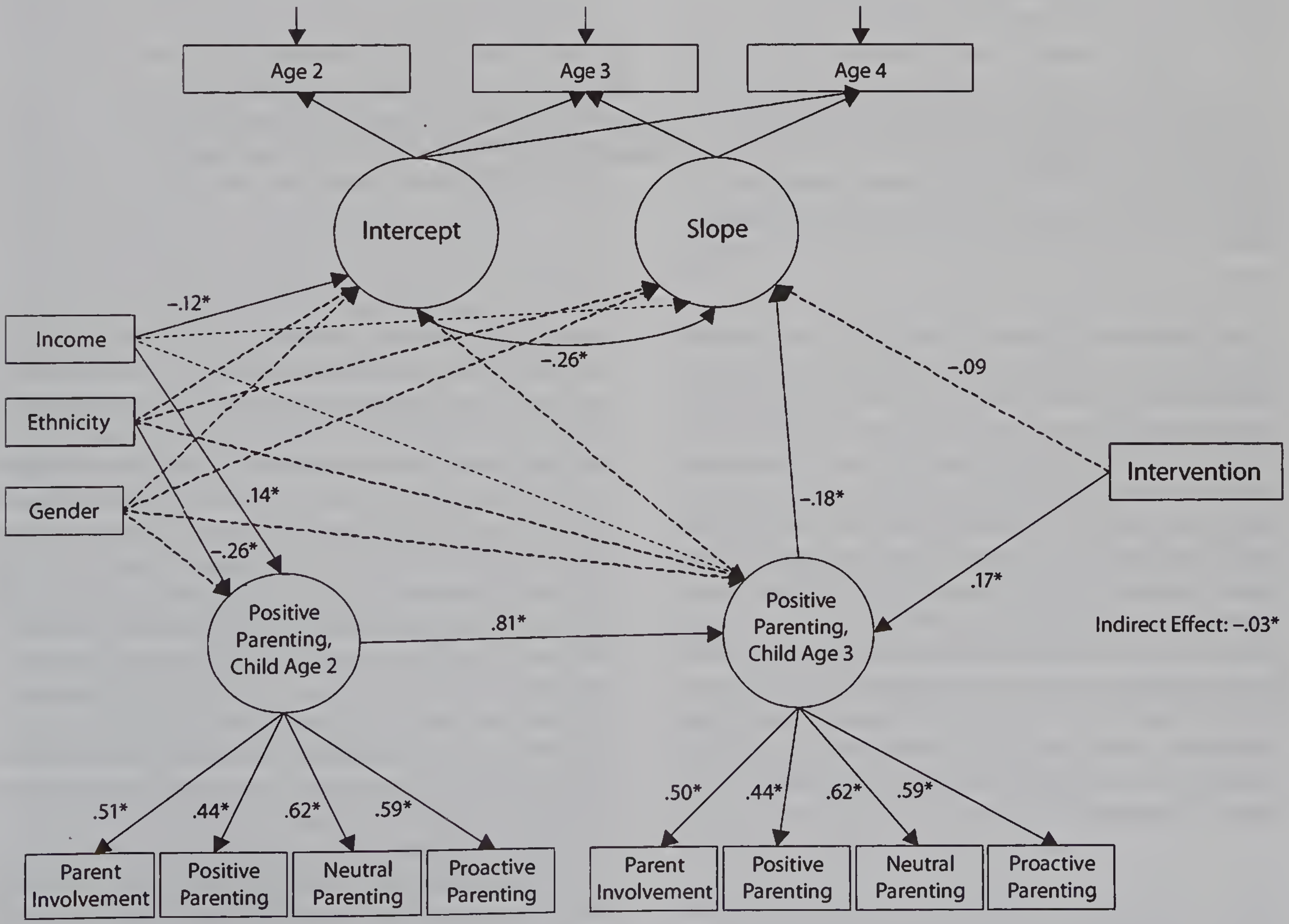

Age 2
Age 3
Age 4
Intercept
Slope
Income
Ethnicity
Gender
−.12*
−.26*
−.09
Intervention
.14*
−.26*
−.18*
.17*
Indirect Effect: −.03*
Positive Parenting, Child Age 2
.81*
Positive Parenting, Child Age 3
.51*
.44*
.62*
.59*
.50*
.44*
.62*
.59*
Parent Involvement
Positive Parenting
Neutral Parenting
Proactive Parenting
Parent Involvement
Positive Parenting
Neutral Parenting
Proactive Parenting

Figure 6. Mediation model: Child Behavior Checklist Externalizing.
$^*p < .05$.

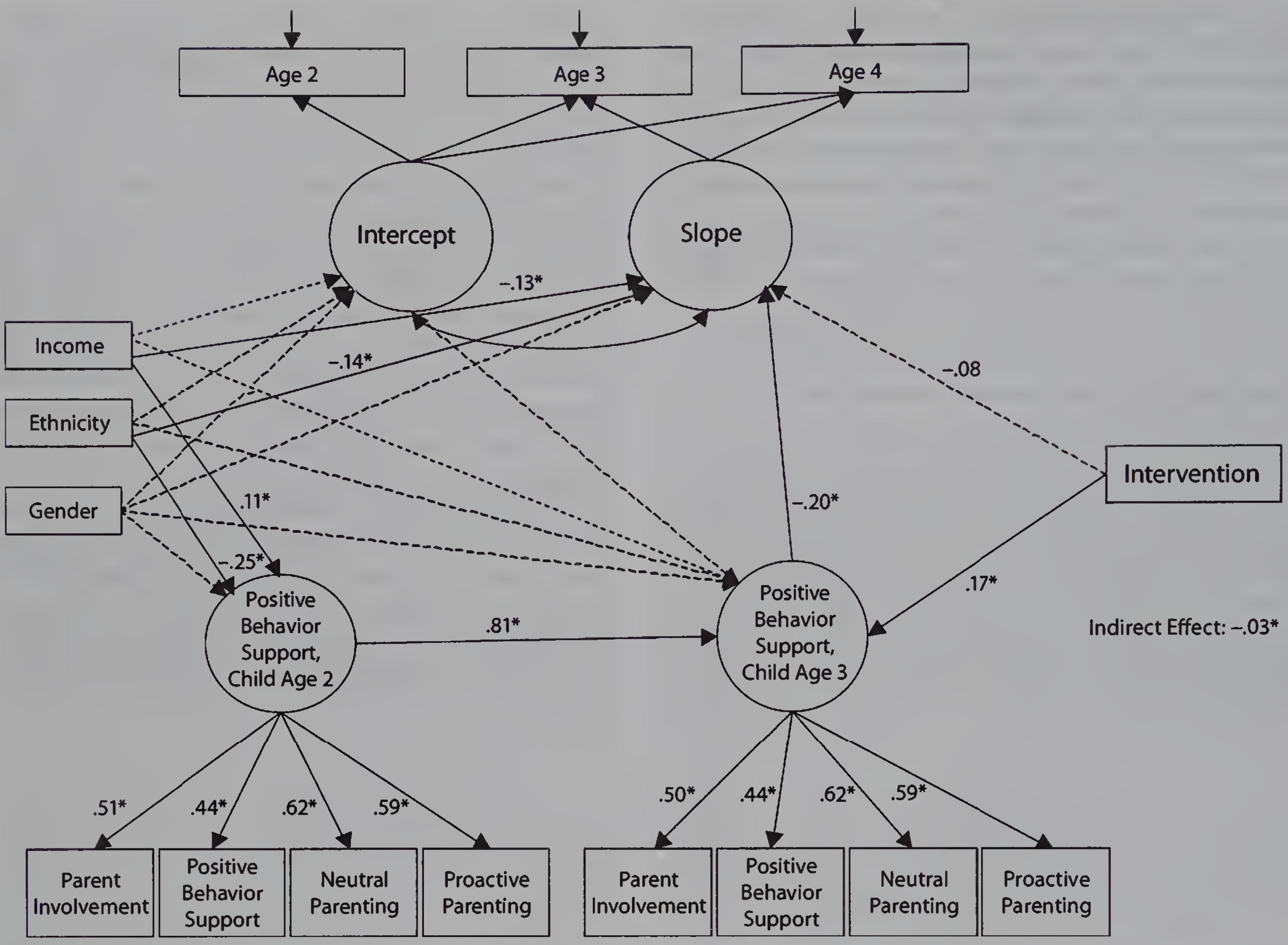

Age 2
Age 3
Age 4
Intercept
Slope
−.13*
Income
Ethnicity
Gender
−.14*
−.08
Intervention
.11*
−.20*
−.25*
.17*
Positive Behavior Support, Child Age 2
.81*
Positive Behavior Support, Child Age 3
Indirect Effect: −.03*
.51*
.44*
.62*
.59*
.50*
.44*
.62*
.59*
Parent Involvement
Positive Behavior Support
Neutral Parenting
Proactive Parenting
Parent Involvement
Positive Behavior Support
Neutral Parenting
Proactive Parenting

Figure 7. Mediation model: Eyberg Child Behavior Inventory.
$^*p < .05$.

As shown in Figure 7, the model for problem behavior from the Eyberg Child Behavior Inventory provided reasonable fit to the data by most indices of model fit, $\chi^2(df = 66) = 89.45$, $p = .03$; CFI = .98; RMSEA = .02; SRMR = .04, and the nonsignificant chi-square may be related to the large sample size. The direct effect of intervention on the problem behavior slope was not significant with positive behavior support in the equation. The family-centered intervention significantly predicted improvements in positive behavior support from child ages 2 to 3. More positive behavior support predicted less growth in problem behavior in early childhood. The indirect effect from intervention to more positive behavior support to less growth in problem behavior was statistically reliable.

Discussion

These data support the hypothesis that a brief, adaptive, and tailored approach to supporting positive behavior support practices can prevent the growth of problem behavior in young children at a critical 2-year period in development (ages 2, 3, and 4 years). Moreover, reduction in growth of early childhood problem behavior was associated with improvement in positive behavior support practices from child ages 2 to 3. The effect sizes for the impact of the intervention were in the small to moderate range ($d = .33$, positive behavior support; $d = .23$, problem behavior) when a one-group intention to treat design analysis was used. Similarly, when a person-centered analysis strategy was used, it was revealed that the intervention effect in child problem behavior was most pronounced among the children who were at highest risk at age 2 and was found to be larger than the effect for the entire sample ($d = .33$).

Three significant points can be considered when evaluating effect sizes in general and those resulting from this study in particular (McCartney & Rosenthal, 2000). First, this is a preventive intervention, and many of the families, although at risk, did not display problem parenting or child behavior and therefore change was not to be expected. The EcoFIT was designed in terms of a health maintenance framework; therefore, continued health is an explicit goal and cannot truly be evaluated until the children are in the next developmental stage, which is school age. Second, it appears that a small to moderate effect size is potentially of some practical significance, if indeed these levels of reduction in problem behavior translate to preventing coercive interaction cycles and escalating patterns of problem behavior that might be expected among the highest risk toddlers (Shaw et al., 2003). In an ongoing study by Shaw and colleagues, indices of parenting assessed at age 2 are being found to predict youth reports of arrest at age 15, for which even modest reductions would be an important outcome (Shaw & Gross, in press). Third, the developmental time span for evaluating the preventive impact of the EcoFIT delivered at this age is truncated, and it is possible that the effect sizes may increase over time, as has been found in several other prevention trials (Connell et al., 2007; DeGarmo, Patterson, & Forgatch, 2004; Ialongo, Poduska, Werthamer, & Kellam, 2001; Olds et al., 1997).

Given these considerations, we propose that the relatively small effect sizes may be of practical significance, although more extensive follow-up of this sample is needed to more formally evaluate this expectation. Given the potential continuity of antisocial behavior over time (Loeber & Dishion, 1983; Olweus, 1979) and the likelihood that early onset leads to multiple forms of problem behavior by adolescence (Dishion & Patterson, 2006), the potential for preventing early-onset and serious problem behavior by adolescents could result in enormous economic savings and benefits to children's lives (Miller, 2004).

It is worth noting that the risk of early-onset antisocial behavior in this sample is considerable. In earlier research, we found that early-onset externalizing problems could be predicted by a combination of maternal depression and active, fearless temperament when children were age 2 (Shaw et al., 2006). We in fact screened the population in this study to include mothers and children with multiple risk factors, including maternal depression and child temperament and behavior problems. Therefore, our ability to effectively engage families, briefly intervene to promote positive behavior support, and reduce growth in problem behavior has three major implications.

First, the EcoFIT model in general and the FCU in particular are promising strategies that can be used as an inexpensive intervention in service settings that involve a large number of families at high risk. The WIC venue is especially suitable because its primary goal is to promote the health of young children ages 0–5. Low-income families in need of nutritional and financial support could also be offered cost-effective parenting services that, for many, could have substantial benefits in the long-term reduction of risk. Although it is certainly true that many of the families would benefit from or require more support than the FCU can provide, the use of assessments to target and focus treatment services significantly reduces the cost of the services and improves the engagement of families at high risk (Dishion & Stormshak, 2007). Communities concerned about promoting the behavioral health of children and ensuring school readiness could use WIC as a venue for empirically supported prevention practices.

Second, this study provides an important addition to a sparse literature on intervention mechanisms (Weersing & Weisz, 2002) by showing that even a brief, individually tailored intervention that typically involves limited opportunity for skills training effectively mediates change in positive behavior support style and skill. Other studies finding this effect have used considerably longer interventions that involved more extensive behavioral skills training (Gardner et al., 2006; Gardner, Hutchings, & Bywater, 2007; Martinez & Forgatch, 2001).

Third, the sample studied was diverse with respect to both gender and ethnicity. As previously stated, a unique advantage of the adaptive, tailored EcoFIT family intervention is that it is flexibly delivered and responsive to a wide variety of families and children. Indeed, we found no significant differences among our study's ethnic groups in terms of engagement or effectiveness of the intervention. Nevertheless, given our experience with diverse families, it is important to consider further improvements in the intervention model's capacity to meet the needs of a wide variety of families. For example, many minority families in the United States deal with challenges such as acculturation stress, discrimination, or simply maintaining cultural strengths and coherence in the face of a dominant, majority community. We are currently developing a broader assessment menu that will enable assessment of the strengths and needs of families of various ethnicities that is relevant to factors such as racial socialization and cultural resilience (Yasui & Dishion, 2007). Presently, however, it is reassuring that the flexibility of the EcoFIT approach and the FCU delivered by a culturally sensitive therapist is helpful to majority and minority families.

This study's findings corroborate results from the broader literature on the effectiveness of preventive interventions aimed at reducing child conduct problems in early childhood among families at high risk

(Baydar, Reid, & Webster-Stratton, 2003; Olds, 2002). The findings also provide additional support for the effectiveness of the FCU in general (Connell et al., 2007; Dishion et al., 2003) and its application to high-risk families during the toddler period. In an earlier study with a smaller sample of toddler-age boys from an urban community, we found that one session of the FCU reduced subsequent child externalizing problems and increased positive behavior support and parental involvement (Gardner et al., 2007; Shaw et al., 2006). The results of our current study suggest that a repeated FCU session engenders more consistent improvements in child behavior and parenting outcomes with a larger and more ethnically diverse sample of boys *and* girls followed from ages 2 to 4.

Finally, the results of this randomized intervention trial provide an important feedback loop to developmental theory. It was originally hypothesized that early support for family management in toddlerhood would reduce the emergence of problem behavior in early childhood. The effect of the intervention based on observations supports three key ideas relevant to developmental psychology. First, early parenting is malleable and responsive to interventions. Concern often arises from longitudinal studies that do not include interventions that associations observed between parenting practices and child behavior are nothing more than genetic effects (e.g., Harris, 1995; Rutter et al., 1997). Our findings suggest that change in parenting practices indeed translates to improved childhood outcomes, as have the findings of other intervention studies that have pointed to a link between changing parenting practices and child and adolescent outcomes (e.g., Dishion, Patterson, & Kavanagh, 1992; Forgatch & DeGarmo, 1999). Second, some change in parenting can result from focusing on caregivers' motivation rather than on direct training, and this was especially true for the highest risk families. This finding suggests a disruption model of poor parenting, in that many parents would engage in more positive parenting practices in the right circumstances but that stress and other family dynamics may disrupt *performance* in childrearing practices (Patterson, 1983). This is an optimistic perspective for developmental and prevention science and for those concerned with improving long-term outcomes for children. Finally, these data suggest that an early focus on positive and proactive parenting practices can prevent the emergence of parent–child coercion dynamics known to be prognostic of more serious antisocial behavior down the line in development. This finding suggests the need to integrate both relational and positive and negative interaction dynamics to further understand the contribution of the caregiving environment in developmental psychopathology.

Limitations

Our study has methodological limitations that merit consideration. First, although we presented evidence to suggest that the FCU is associated with improvements in child problem behavior and positive behavior support, effect sizes, albeit meaningful from a public health perspective, were relatively modest. There are two reasons for the small effect sizes. One, although families in the study were at high risk, many of them were actually doing well when evaluated with standardized measures of child problem behavior and direct observation. Therefore, a change was not expected or induced. Two, we suspect that for those families with more serious difficulties that do not change, two patterns of core issues are basic to the family disruption. One is the caregivers' affective adjustment, usually depression as a function of having several young children under conditions of poverty. We considered the effects of the intervention on maternal depression in another analysis of these data and indeed found reduced levels of maternal depression as a function of the intervention (Shaw, Connell, Dishion, Wilson, & Gardner, 2008). The other is a history of maternal trauma that underlies both the depression and the poverty conditions of the family. Psychosocial interventions such as EcoFIT will have limited impact in some contexts and conditions, and these conditions require attention at the broader level of the community and the nation. Nevertheless, the EcoFIT model can be helpful for some caregivers in affective distress (Shaw et al., 2008), and we plan to further develop this component of the intervention model following principles derived from evidence-based practices that target depression, dissociation, and anxiety secondary to difficult circumstances and history.

A second limitation of the study is that for this age, we had only one reporting agent to address the clinical significance of the child's problem behavior. Because participation in the intervention group might have biased later maternal reports of child behavior, it would have been more optimal to have had a second informant of child behavior. Relevant data will be forthcoming from teachers in the coming years. How well a child adapts to the public school environment is a key test of the overall benefit of early intervention because this critical transition has significant long-term consequences for the child's future and adjustment (Kellam & Van Horn, 1997). Continued follow-up of our study sample will shed light on the endurance of intervention gains seen in the home context and whether such effects are evident at school.

Clearly, many traditional models of mental health intervention are wanting with respect to the most vulnerable families and children. Consistent with insights shared by Kazdin (1987) and others (Dishion & Patterson, 1993; Patterson, Dishion, & Chamberlain, 1993), a health maintenance model is likely more appropriate than a "medical model." That is, continued support helps vulnerable families and children continue to maintain healthy parenting practices and relationship patterns that promote children's adjustment through adolescence and into young adulthood. Thus, we now cast EcoFIT as a health maintenance model designed for implementation in contexts that facilitate regular and repeated exposure and influence on children and families and less for freestanding mental health clinics that do not have ties to institutions such as public schools (see Dishion & Stormshak, 2007).

Our current findings endorse previous evidence that long-term improvements in young children's problem behavior can be achieved with a brief family-based intervention for toddlers. Our study also showed that improving positive behavior support practices mediated positive child behavior change. Reduction in child problem behaviors was achieved using an existing, nationally available service delivery setting for low-income children who are at risk for early-starting pathways of externalizing problem behavior and whose families do not typically use mental health services, especially this early in their child's development. Future follow-up of this study's cohort should clarify questions regarding the intervention's endurance, potential changes in effect sizes, and the benefits of the changes to the child's adaptation to the school environment.

References

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms & profiles: Child Behavior Checklist for Ages 6–18, Teacher's Report Form, & Youth Self-Report*. Burlington: University of Vermont, Research Center for Children, Youth & Families.

Baldwin, D. A. (1995). Understanding the link between joint attention and language. In P. J. Dunham & C. Moore (Eds.), *Joint attention: Its origins and roles in development* (pp. 131–158). Hillsdale, NJ: Erlbaum.

Baldwin, D. A., Bill, B., Desjardins, R. N., Irwin, J. M., & Markman, E. M. (1996). Infants' reliance on social criterion for establishing word–object relations. *Journal of Child Development*, *67*, 3135–3153.

Baydar, N., Reid, M. J., & Webster-Stratton, C. (2003). The role of mental health factors and program engagement in the effectiveness of a preventive parenting program for Head Start mothers. *Child Development*, *74*, 1433–1453.

Bradley, R. H., Corwyn, R. F., McAdoo, H. P., & Garcia-Coll, C. (2001). The home environments of children in the United States. Part I: Variations by age, ethnicity, and poverty status. *Child Development*, *72*, 1844–1867.

Brinkmeyer, M. Y., & Eyberg, S. M. (2003). Parent–child interaction therapy for oppositional children. In A. E. Kazdin & J. R. Weisz (Eds.), *Evidence-based psychotherapies for children and adolescents* (pp. 204–223). New York: Guilford.

Brody, G. H., Murry, V. M., Gerrard, M., Gibbons, F. X., Molgaard, V., & McNair, L., et al. (2004). The strong African American families program: Translating research into prevention programming. *Child Development*, *75*, 900–917.

Brook, J. S., Whiteman, M., Cohen, P., & Tanaka, J. S. (1992). Childhood precursors of adolescent drug use: A longitudinal analysis. *Genetic, Social, and General Psychology Monographs*, *118*, 195–213.

Brook, J. S., Whiteman, M., Gordon, A. S., & Cohen, P. (1986). Dynamics of childhood and adolescent personality traits and adolescent drug use. *Developmental Psychology*, *22*, 403–414.

Bullock, B. M., & Forgatch, M. S. (2005). Mothers in transition: Model-based strategies for effective parenting. In W. M. Pinsof & J. L. Lebow (Eds.), *Family psychology: The art of the science* (pp. 349–371). New York: Oxford University Press.

Campbell, S. B., Pierce, E. W., Moore, G., Marakovitz, S., & Newby, K. (1996). Boys' externalizing problems at elementary school age: Pathways from early behavior problems, maternal control, and family stress. *Development and Psychopathology*, *8*, 701–719.

Campbell, S. B., Shaw, D. S., & Gilliom, M. (2000). Early externalizing behavior problems: Toddlers and preschoolers at risk for later maladjustment. *Development and Psychopathology*, *12*, 467–488.

Capaldi, D. M., Crosby, L., & Stoolmiller, M. (1996). Predicting the timing of first sexual intercourse for at-risk adolescent males. *Child Development*, *67*, 344–359.

Caspi, A., McClay, J., Moffitt, T., Mill, J., Craig, I. W., & Taylor, A., et al. (2002). Role of genotype in the cycle of violence in maltreated children. *Science*, *297*, 851–854.

Caspi, A., Moffitt, T. E., Newman, D. L., & Silva, P. A. (1998). Behavioral observations at age 3 years predict adult psychiatric disorders: Longitudinal evidence from a birth cohort. In M. E. Hertzig & E. A. Ellen (Eds.), *Annual progress in child psychiatry and child development* (pp. 319–331). Philadelphia: Brunner/Mazel.

Conduct Problems Prevention Research Group. (2002). Evaluation of the first 3 years of the Fast Track Prevention Trial with children at high risk for adolescent conduct problems. *Journal of Abnormal Child Psychology*, *30*, 19–35.

Connell, A., Dishion, T., Yasui, M., & Kavanagh, K. (2007). An adaptive approach to family intervention: Linking engagement in family-centered intervention to reductions in adolescent problem behavior. *Journal of Consulting and Clinical Psychology, 75*, 568–579.

DeGarmo, D. S., Patterson, G. R., & Forgatch, M. S. (2004). How do outcomes in a specified parent training intervention maintain or wane over time? *Prevention Science, 5*(2), 73–89.

DiBhion, T. J., Patterson, G. R., & Kavanagh, K. (1992). An experimental test of the coercion model: Linking theory, measurement, and intervention. In J. McCord & R. Trembley (Eds.), *The interaction of theory and practice: Experimental studies of interventions* (pp. 253–282). New York: Guilford.

Dishion, T. J. (1990). The family ecology of boys' peer relations in middle childhood. *Child Development, 61*, 874–892.

Dishion, T. J., Duncan, T. E., Eddy, J. M., Fagot, B. I., & Fetrow, R. (1994). The world of parents and peers: Coercive exchanges and children's social adaptation. *Social Development, 3*, 255–268.

Dishion, T. J., Gardner, K., Patterson, G. R., Reid, J. B., & Thibodeaux, S. (1983). *The Family Process Code: A multidimensional system for observing family interaction*. Unpublished coding manual. Available from Oregon Social Learning Center, 160 East 4th Avenue, Eugene, OR 97401-2426.

Dishion, T. J., & Kavanagh, K. (2003). *Intervening in adolescent problem behavior: A family-centered approach*. New York: Guilford.

Dishion, T. J., Nelson, S. N., & Bullock, B. M. (2004). Premature adolescent autonomy: Parent disengagement and deviant peer process in the amplification of problem behavior. *Journal on Adolescence, 27*, 515–530.

Dishion, T. J., Nelson, S. E., & Kavanagh, K. (2003). The Family Check-Up for high-risk adolescents: Preventing early-onset substance use by parent monitoring. *Behavior Therapy, 34*, 553–571.

Dishion, T. J., & Owen, L. D. (2002). A longitudinal analysis of friendships and substance use: Bidirectional influence from adolescence to adulthood. *Developmental Psychology, 28*, 480–491.

Dishion, T. J., & Patterson, G. R. (1992). Age effects in parent training outcome. *Behavior Therapy, 23*, 719–729.

Dishion, T. J., & Patterson, G. R. (1993). Childhood screening for early adolescent problem behavior: A multiple gating strategy. In M. Singer, L. Singer, & T. M. Anglin (Eds.), *Handbook for screening adolescents at psychosocial risk* (pp. 375–399). New York: Lexington Books.

Dishion, T. J., & Patterson, G. R. (2006). The development and ecology of antisocial behavior in children and adolescents. In D. Cicchetti & D. J. Cohen (Eds.), *Developmental psychopathology: Vol. 3. Risk, disorder, and adaptation* (pp. 503–541). New York: Wiley.

Dishion, T. J., & Stormshak, E. (2007). *Intervening in children's lives: An ecological, family-centered approach to mental health care*. Washington, DC: APA Books.

Dumas, J. E., Lemay, P., & Dauwalder, J. (2001). Dynamic analyses of mother–child interactions in functional and dysfunctional dyads: A synergetic approach. *Journal of Abnormal Child Psychology, 29*, 317–329.

Eddy, J. M., & Chamberlain, P. (2000). Family management and deviant peer association as mediators of the impact of treatment condition on youth antisocial behavior. *Journal of Consulting and Clinical Psychology, 5*, 857–863.

Falloon, I. R., Boyd, J. L., McGill, C. W., Williamson, M., Razani, J., Moss, H. B., et al. (1985). Family management in the prevention of morbidity of schizophrenia. Clinical outcome of a two-year longitudinal study. *Archives of General Psychiatry, 42*, 887–896.

Forgatch, M. S., Bullock, B. M., & Patterson, G. R. (2004). From theory to practice: Increasing effective parenting through role play. The Oregon Model of Parent Management Training (PMTO). In H. Steiner, K. Chang, J. Lock, & J. Wilson (Eds.), *Handbook of mental health interventions in children and adolescents: An integrated development approach* (pp. 782–813). San Francisco: Jossey-Bass.

Forgatch, M. S., & DeGarmo, D. S. (1999). Parenting through change: An effective prevention program for single mothers. *Journal of Consulting & Clinical Psychology, 67*, 711–724.

Forgatch, M. S., Patterson, G. R., & DeGarmo, D. S. (2005). Evaluating fidelity: Predictive validity for a measure of competent adherence to the Oregon Model of Parent Management Training. *Behavior Therapy, 36*, 3–13.

Forgatch, M., & Toobert, D. J. (1979). A cost-effective parent training program for use with normal preschool children. *Journal of Pediatric Psychology, 4*(2), 129–145.

French, D. C., & Dishion, T. J. (2003). Predictors of early initiation of sexual intercourse among high-risk adolescents. *Journal of Early Adolescence, 23*, 295–315.

Gardner, F. E. M. (1994). The quality of joint activity between mothers and their children with conduct problems. *Journal of Child Psychology and Psychiatry, 35*, 935–949.

Gardner, F., Burton, J., & Klimes, I. (2006). Randomized controlled trial of a parenting intervention in the voluntary sector for reducing child conduct problems: Outcomes and mechanisms of change. *Journal of Child Psychology & Psychiatry, 47*, 1123–1132.

Gardner, F., Hutchings, J., & Bywater, T. (2007). *Moderators and mediators of intervention effects in a randomised trial of a parenting intervention in Sure Start services*. Manuscript submitted for publication.

Gardner, F., Shaw, D., Dishion, T. J., Burton, J., & Supplee, L. (2007). Randomized prevention trial for early conduct problems: Effects on proactive parenting and links to toddler disruptive behavior. *Journal of Family Psychology, 21*, 398–406.

Gardner, F., Sonuga-Barke, E., & Sayal, K. (1999). Parents anticipating misbehavior: An observational study of strategies parents use to prevent conflict with behavior problem children. *Journal of Child Psychology and Psychiatry, 40*, 1185–1196.

Gardner, F., Ward, S., Burton, J., & Wilson, C. (2003). Joint play and the early development of conduct problems in children: A longitudinal observational study of preschoolers. *Social Development, 12*, 361–379.

Glik, D. C., Greaves, P. E., Kronenfeld, J. J., & Jackson, K. L. (1993). Safety hazards in households with young children. *Journal of Pediatric Psychology, 18*(1), 115–131.

Harris, J. R. (1995). Where is the child's environment? A group socialization theory of development. *Psychological Review, 102*, 458–489.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore: Brookes.

Hawkins, J. D., Lishner, D. M., Catalano, R. F., & Howard, M. O. (1986). Childhood predictors of adolescent substance abuse: Toward an empirically grounded theory. *Journal of Children in Contemporary Society, 18*, 11–47.

Henggeler, S., Schoenwald, S., Borduin, C., Rowland, M., & Cunningham, P. (1998). *Multisystemic treatment of antisocial behavior in children and adolescents*. New York: Guilford.

Hoagwood, K., & Koretz, D. (1996). Embedding prevention services within systems of care: Strengthening the nexus for children. *Applied and Preventative Psychology, 5*, 225–234.

Horner, R. H., & Carr, E. G. (1997). Behavioral support for students with severe disabilities: Functional assessment intervention. *Journal of Special Education, 31*, 84–104.

Hutchings, J., Bywater, T., Daley, D., Gardner, F., Jones, K., Eames, C., et al. (2007). Pragmatic randomized controlled trial of a parenting intervention in "Sure Start" services for children at risk of developing conduct disorder. *British Medical Journal, 334*, 678–685.

Ialongo, N., Poduska, J., Werthamer, L., & Kellam, S. (2001). The distal impact of two first-grade preventive interventions on conduct problems and disorder in early adolescence. *Journal of Emotional and Behavioral Disorders, 9*, 146–160.

Ingoldsby, E., & Shaw, D. S. (2002). Neighborhood contextual factors and the onset and progression of early-starting antisocial pathways. *Clinical Child and Family Psychology Review, 5*(1), 21–55.

Ingoldsby, E., Shaw, D. S., & Garcia, M. (2001). Intrafamilial conflict in relation to boys' adjustment at school. *Development and Psychopathology, 13*, 35–52.

Jabson, J. M., Dishion, T. J., Gardner, F. E. M., & Burton, J. (2004). *Relationship Process Code v-2.0. training manual: A system for coding relationship interactions*. Unpublished coding manual. Available from the Child and Family Center, 160 East 4th Avenue, Eugene, OR 97401-2426.

Jones, L. M. (2004). The Family Check-Up for families of high-risk preschoolers: The moderating effect of children's temperament. *.Dissertation Abstracts International, 64*(8-B), 4083.

Kazdin, A. E. (1987). Treatment of antisocial behavior in children: Current status and future directions. *Psychological Bulletin, 102*, 187–203.

Kazdin, A. E. (2003). Problem-solving skills training and parent management training for conduct disorder. In A. E. Kazdin & J. R. Weisz (Eds.), *Evidence-based psychotherapies for children and adolescents* (pp.. 241–262). New York: Guilford.

Kellam, S. G., & Van Horn, Y. V. (1997). Life course development, community epidemiology, and preventive trials: A scientific structure for prevention research. *American Journal of Community Psychology, 25*, 177–188.

Liddle, H. A. (1999). Theory in a family-based therapy for adolescent drug abuse. *Journal of Clinical Child Psychology, 28*, 521–532.

Loeber, R., & Dishion, T. (1983). Early predictors of male delinquency: A review. *Psychological Bulletin, 94*, 68–99.

Lyons-Ruth, K., Zeanah, C. H., & Benoit, D. (1996). Disorder and risk for disorder during infancy and toddlerhood. In E. J. Mash & R. A. Barkley (Eds.), *Child psychopathology* (pp. 457–491). New York: Guilford.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods, 7*(1), 83–104.

MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research, 39*(1), 99–128.

Martinez, C., & Forgatch, M. (2001). Preventing problems with boys' noncompliance: Effects of a parent training intervention for divorcing mothers. *Journal of Consulting and Clinical Psychology, 69*, 416–428.

McCartney, K., & Rosenthal, R. (2000). Effect size, practical importance, and social policy for children. *Child Development, 71*, 173–180.

Miller, T. R. (2004). The social costs of adolescent problem behavior. In A. Biglan, P. A. Brennan, S. L. Foster, & H. D. Holder (Eds.), *Helping adolescents at risk: Prevention of multiple problem behaviors* (pp. 31–56). New York: Guilford.

Miller, W. R., & Rollnick, S. (2002). *Motivational interviewing: Preparing people for change* (2nd ed.). New York: Guilford.

Muthén, B. O., & Muthén, L. (2000). Integrating person-centered and variable-centered analysis: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research, 24*, 882–891.

Muthén, L. K., & Muthén, B. O. (2004). *Mplus user's guide* (3rd ed.). Los Angeles: Author.

Muthén, L. K., & Muthén, B. O. (2007). *Mplus user's guide* (5th ed.). Los Angeles: Author.

Nylund, K., Asparouhov, T., & Muthén, B. (2006). *Deciding on the number of classes in mixture modeling: A montecarlo simulation study.* Manuscript submitted for publication.

Olds, D. L. (2002). Prenatal and infancy home visiting by nurses: From randomized trials to community replication. *Prevention Science, 3*(3), 153–172.

Olds, D. L., Eckenrode, J., Henderson, C. R., Jr., Kitzman, H., Powers, J., & Cole, R., et al. (1997). Long-term effects of home visitation on maternal life course and child abuse and neglect. Fifteen-year follow-up of a randomized trial. *Journal of the American Medical Association, 278*, 637–643.

Olds, D. L., Hill, P., Robinson, J., Song, N., & Little, C. (2000). Update on home visiting for pregnant women and parents of young children. *Current Problems in Pediatrics, 30*(4), 110–114.

O'Leary, C. C. (2000). *A brief family intervention for high-risk preschool-aged children*. Unpublished doctoral dissertation, University of Oregon, Eugene.

Olweus, D. (1979). Stability of aggressive reaction patterns in males: A review. *Psychological Bulletin, 86*, 852–875.

Patterson, G. R. (1982). *A social learning approach: III. Coercive family process*. Eugene, OR: Castalia.

Patterson, G. R. (1983). Stress: A change agent for family process. In N. Garrezy & M. Rutter (Eds.), *Stress, coping, and development in children* (pp. 235–264). New York: McGraw-Hill.

Patterson, G. R., & Dishion, T. J. (1988). Multilevel family process models: Traits, interactions, and relationships. In R. A. Hinde & J. Stevenson-Hinde (Eds.), *Relationships and families: Mutual influences* (pp. 283–310). Oxford, UK: Claredon.

Patterson, G. R., Dishion, T. J., & Chamberlain, P. (1993). Outcomes and methodological issues relating to treatment of antisocial children. In T. R. Giles (Ed.), *Effective psychotherapy: A handbook of comparative research* (pp. 43–88). New York: Plenum.

Patterson, G. R., Dishion, T. J., & Yoerger, K. (2000). Adolescent growth in new forms of problem behavior: Macro- and micro-peer dynamics. *Prevention Science, 1*, 3–13.

Patterson, G. R., Reid, J. B., & Dishion, T. J. (1992). *Antisocial boys*. Eugene, OR: Castalia.

Pettit, G. S., & Bates, J. E. (1989). Family interaction patterns and children's behavior problems from infancy to 4 years. *Developmental Psychology, 25*, 413–420.

Pianta, R. C., Steinberg, M. S., & Rollins, K. B. (1995). The first two years of school: Teacher–child relationships and deflections in children's classroom adjustment. *Development and Psychopathology, 7*, 295–312.

Poulin, F., & Boivin, M. (2000). The formation and development of friendship in childhood: The role of proactive and reactive aggression. *Developmental Psychology, 36*, 233–240.

Radloff, L. S. (1977). The CED-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*, 385–401.

Robinson, E. A., Eyberg, S. M., & Ross, A. W. (1980). The standardization of an inventory of child conduct problem behaviors. *Journal of Clinical Child Psychology, 9*, 22–28.

Rutter, M., Dunn, J., Plomin, R., Simonoff, E., Pickles, A., Maughan, B., et al. (1997). Integrating nature and nurture: Implications of person–environment correlations and interactions for developmental psychopathology. *Development and Psychopathology, 9*, 335–364.

Sameroff, A. J., & Fiese, B. H. (1987). *Conceptual issues in prevention*. Unpublished manuscript.

Sanders, M. R. (1999). Triple P-Positive Parenting Program: Toward an empirically validated multilevel parenting and family support strategy for the prevention of behavior and emotional problems in children. *Clinical Child and Family Psychology Review, 2*, 71–90.

Shaw, D. S., Bell, R. Q., & Gilliom, M. (2000). A truly early-starter model of antisocial behavior revisited. *Clinical Child and Family Psychology Review, 3*(3), 155–172.

Shaw, D., Connell, A., Dishion, T., Wilson, M., & Gardner, F. (2008). *Improvements in maternal depression as a mediator of intervention effects on early childhood problem behavior and affect*. Manuscript under review.

Shaw, D. S., Dishion, T. J., Supplee, L., Gardner, F., & Arnds, K. (2006). A family-centered approach to the prevention of early-onset antisocial behavior: Two-year effects of the Family Check-Up in early childhood. *Journal of Consulting and Clinical Psychology, 74*, 1–9.

Shaw, D. S., Gilliom, M., Ingoldsby, E. M., & Nagin, D. (2003). Trajectories leading to school-age conduct problems. *Developmental Psychology, 39*, 189–200.

Shaw, D. S., & Gross, H. (in press). Early childhood and the development of delinquency: What we have learned from recent longitudinal research. In A. Lieberman (Ed.), *The yield of recent longitudinal studies of crime and delinquency.* New York: Springer.

Shaw, D. S., Owens, E. B., Vondra, J. I., Keenan, K., & Winslow, E. B. (1996). Early risk factors and pathways in the development of early disruptive behavior problems. *Development and Psychopathology, 8*, 679–699.

Snyder, J., Edwards, P., McGraw, K., Kilgore, K., & Holten, A. (1993). *Escalation and reinforcement in family conflict: Development origins of physical aggression*. Wichita: Kansas State University.

Spoth, R., Redmond, C., & Shin, C. (1998). Direct and indirect latent-variable parenting outcomes of two universal family-focused preventative interactions: Extending a public health oriented research base. *Journal of Consulting and Clinical Psychology, 66*, 385–399.

Sugai, G., Horner, R. H., & Sprague, J. R. (1999). Functional-assessment-based behavior support planning: Research to practice to research. *Behavior Disorders, 24*, 253–257.

Supplee, L. H., Unikel, E., & Shaw, D. S. (2007). Physical environmental adversity and the protective role of maternal monitoring in relation to early child conduct problems. *Journal of Applied Developmental Psychology, 28*, 166–183.

Vicary, J. R., & Lerner, J. V. (1983). Longitudinal predictors of drug use: Analyses from the New York Longitudinal Study. *Journal of Drug Education, 13*, 275–285.

Webster-Stratton, C. (1985). Comparisons of behavior transactions between conduct-disordered children and their mothers in the clinic and at home. *Journal of Abnormal Child Psychology, 13*, 169–183.

Webster-Stratton, C. (1990). Long-term follow-up of families with young conduct problem children: From preschool to grade school. *Journal of Clinical Child Psychology, 19*, 144–149.

Webster-Stratton, C., & Taylor, T. (2001). Nipping early risk factors in the bud: Preventing substance abuse, delinquency, and violence in adolescence through interventions targeted at young children (0 to 8 years). *Prevention Science, 2*(3), 165–192.

Weersing, V., & Weisz, J. (2002). Mechanisms of action in youth psychotherapy. *Journal of Child Psychology and Psychiatry, 43*, 3–29.

Yasui, M., & Dishion, T. J. (2007). The ethnic context of child and adolescent problem behavior: Implications for child and family interventions. *Clinical Child and Family Psychology Review, 10*(2), 137–179.

Child Development, September/October 2008, Volume 79, Number 5, Pages 1415 – 1431

Maternal Socialization of Positive Affect: The Impact of Invalidation on Adolescent Emotion Regulation and Depressive Symptomatology

Marie B. H. Yap and Nicholas B. Allen
ORYGEN Research Centre and University of Melbourne

Cecile D. Ladouceur
University of Pittsburgh

This study examined the relations among maternal socialization of positive affect (PA), adolescent emotion regulation (ER), and adolescent depressive symptoms. Two hundred early adolescents, 11 – 13 years old, provided self-reports of ER strategies and depressive symptomatology; their mothers provided self-reports of socialization responses to adolescent PA. One hundred and sixty-three mother – adolescent dyads participated in 2 interaction tasks. Adolescents whose mothers responded in an invalidating or "dampening" manner toward their PA displayed more emotionally dysregulated behaviors and reported using maladaptive ER strategies more frequently. Adolescents whose mothers dampened their PA more frequently during mother – adolescent interactions, and girls whose mothers reported invalidating their PA, reported more depressive symptoms. Adolescent use of maladaptive ER strategies mediated the association between maternal invalidation of PA and early adolescents' concurrent depressive symptoms.

The ability to successfully regulate affective states is central to mental health, and indeed, it has been noted that emotion dysregulation figures prominently in mental illness (Gross & Munoz, 1995). This may be particularly true of depressive disorders, as they are characterized by deficits of positive affect (PA) and/or an excess of negative affect (NA; Gross & Levenson, 1997). Because depressive disorders involve dysfunction of both the major affective systems (i.e., both aversive and appetitive; P. M. Cole & Kaslow, 1988; Gross & Munoz, 1995), failures or abnormalities in emotion regulation (ER) processes are likely to play a particularly critical role in their etiology.

There is much debate in the field regarding the definition of ER (e.g., P. M. Cole, Martin, & Dennis, 2004), and some researchers assert that ER and emotion generation/reactivity are largely inseparable processes (Campos, Frankel, & Camras, 2004). In this article, we acknowledge the critical, and to some extent unresolved, points raised in this debate and hence conceptualize ER more broadly to incorporate processes of emotional reactivity and its effortful regulation. Specifically, ER refers to the internal and transactional processes through which individuals alter or maintain an affective state, often in the service of accomplishing their prioritized, situation-specific goals (Diamond & Aspinwall, 2003; Gross, 1999). ER can occur at any level of the emotion process, at any time the emotion is activated, and can be evident before an emotion is manifested (Campos et al., 2004). Difficulties with altering an undesirable affective state or maintaining a desirable state will henceforth be referred to as "emotion dysregulation." In particular, dysregulation of NA can involve an individual being (too) easily drawn into a negative affective state (when directly or indirectly provoked), becoming "stuck" in that undesired state (i.e., maintains the NA for longer than desired) and/or using strategies that are ineffective in altering the affect as desired. Likewise, PA dysregulation can involve less frequent PA (e.g., less easily drawn into that desirable state) and an inability to maintain one's positive affective state for a longer duration.

For example, Tomarken and Keener (1998) predicted that depressed individuals should display longer maintenance (i.e., greater "temporal continuity") of NA and poorer ability to maintain PA. Consistent with this prediction, depressed adolescents were found to maintain depressive affective states for a longer duration compared to their nondepressed peers during parent – adolescent interactions (Sheeber, Allen, Davis, & Sorensen, 2000). Relatedly, emerging evidence suggests that emotion dysregulation in at-risk or depressed individuals may be linked to deficits in the strategies they use to

---

The research described in this article was partially supported by Program Grant 350241 from the Australian National Health and Medical Research Council and a philanthropic grant from the Colonial Foundation.

Correspondence concerning this article should be addressed to Nicholas B. Allen, ORYGEN Research Centre, and Department of Psychology, University of Melbourne, Melbourne, Victoria 3010, Australia. Electronic mail may be sent to nba@unimelb.edu.au.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7905-0013

respond to negative emotional states (e.g., Garber, Braafladt, & Weiss, 1995; Garber, Braafladt, & Zeman, 1991). One implication of these findings is that a better understanding of how ER skills develop will be valuable for the design of targeted prevention and early intervention strategies for young people with ER difficulties and at risk for depression, especially given that adolescence constitutes a time of particular risk for the onset of depressive disorders (e.g., Kessler, Avenevoli, & Merikangas, 2001; Lewinsohn, Hops, Roberts, Seeley, & Andrews, 1993).

Research indicates that the family caregiving environment plays a crucial role in children's development of ER skills (e.g., Cassidy, 1994; Field, 1994). It has been postulated that caregivers provide "scaffolding" for a child's emotional development (Astington, 1996; Southam-Gerow & Kendall, 2002), and individuals' relationships with their parents set a developmental context for their ER in adolescence and young adulthood (Morris, Silk, Steinberg, Myers, & Robinson, 2007). Evidence suggests that parental behaviors that are low in support and high in conflict are associated with adolescents' regulation of dysphoric affect and experience of depressive symptomatology (e.g., J. P. Allen, Hauser, Eickholt, Bell, & O'Connor, 1994; Sheeber, Hops, Alpert, Davis, & Andrews, 1997; Sheeber, Hops, Andrews, Alpert, & Davis, 1998). Indeed, the family remains a key influence in the child's ER (Sheeber et al., 2000) and mental health (e.g., Gottman, Katz, & Hooven, 1997; Ramsden & Hubbard, 2002) even during adolescence. As Calkins and Bell (1999) have argued, developmental shifts during the transition to adolescence set an important context for parental involvement in the regulation of emotions. To date, however, research on ER during the adolescent period has been largely neglected, particularly in youth at risk of depression (for exceptions, see Kobak, Cole, Ferenz-Gillies, Fleming, & Gamble, 1993; Sheeber et al., 2000; Silk, Steinberg, & Morris, 2003). Moreover, there is evidence that the family remains a vital source of support during adolescence and that family relations have a stronger association with adolescents' depressive symptomatology than peer relations (e.g., Barrera & Garrison-Jones, 1992; McFarlane, Bellissimo, & Norman, 1995). This study aims to examine the role of parental socialization (as one form of family processes) in the development of ER and depressive symptoms in early adolescence.

Parental Socialization of ER

Though the majority of research on emotion socialization has focused on very young children (e.g., P. M. Cole & Kaslow, 1988; Fox, 1994; also see review by Keenan, 2000), preliminary evidence suggests that older children and adolescents may learn to regulate their emotional arousal through parents' responses to their emotional expressions or behavior (e.g., Eisenberg, Fabes, & Murphy, 1996; Eisenberg et al., 1999; Klimes-Dougan et al., 2007).

Research to date on parental responses to child NA indicates that parental socialization characterized by acceptance, coaching, and problem solving are positively associated with children's adaptive responses to NA and inversely associated with depressive symptomatology (Eisenberg et al., 1996; Garber & Dodge, 1991; Gottman, Katz, & Hooven, 1996). Nonetheless, parental socialization of children's PA, and its effect on adolescent emotional functioning and mental health, is still a largely neglected issue in this field. Moreover, research examining the role of parents in emotion socialization after early childhood, especially leading into the adolescence period, has also been scarce, despite this being a critical time for the onset of depression (Kessler et al., 2001).

The pertinence of examining the effects of parental socialization of PA (as distinct from NA) is directly and indirectly implicated by evidence from two fields of research. The first set of evidence comes from the broader parenting literature. Research with young children has demonstrated that parents who frequently discuss emotions (positive and negative), including explanation of the causes and consequences of emotions and reactions to children's expressed emotion, tend to have children with better emotional skills in early childhood (Denham & Kochanoff, 2002; Denham, Mitchell-Copeland, Strandberg, Auerbach, & Blair, 1997; Eisenberg & Fabes, 1994; Halberstadt, Crisp, & Eaton, 1999), which is in turn associated with better ER skills (Fredrickson, 1998; Southam-Gerow & Kendall, 2002). Moreover, parents' positive attentiveness toward their preschoolers' positive and negative affective displays was related to children's emotional and social competence (Denham et al., 1997). However, given the documented changes that occur in the dynamics of parent–child relationships as children approach adolescence (see review by Yap, Allen, & Sheeber, 2007), more research in older childhood and early adolescence is needed. Furthermore, for the purpose of delineating the direct influence of parents on *adolescent ER*, an examination of specific parenting behaviors in response to their adolescent child's behaviors may be more informative. Besides, specific parenting behaviors with adolescents may also be able to inform the focus of interventions for adolescent psychopathology more effectively.

The second set of evidence is presented in a recent review on the relevance of neural systems of PA to the understanding of child and adolescent depression (Forbes & Dahl, 2005). Citing neurobehavioral evidence, namely, in the neural substrates of reward processing, the Forbes and Dahl (2005) distinguish between PA and NA, cogently arguing for the importance of examining the role of PA in youth depression. In particular, the function and interrelation of neural circuits involved in reward processing, including the striatum, orbitofrontal cortex, and amygdala, have specific associations with behavioral manifestations and are indicative of the independence of PA from NA. An important implication of this is the need to examine how children's ability to regulate their positive affective system can be disrupted, and the role of parents (protective or otherwise) in this process.

Of note, although PA is by definition pleasant and desirable, it can be considered inappropriate and/or excessive in certain contexts, especially by parents toward their child. Whether this is due to social norms or expectations, or parental intolerance or poor ER (e.g., in parental depression), parents sometimes respond to child PA in an invalidating manner, by restricting (e.g., instructing child to quiet down), punishing (e.g., reprimand or angry response), or dampening (dysphoric or still-faced response) their child's PA expression. Children who receive such socialization responses to their PA over time may learn to suppress their PA (or at least its expression) and may fail to develop adaptive skills in regulating their PA (Denham et al., 1997). These in turn may contribute to anhedonic tendencies that are purportedly a unique feature of depressive disorders (Clark & Watson, 1991). The overall aim of this study was to investigate whether such invalidating parental socialization of PA is associated with early adolescents' ER and depressive symptoms and whether the socialization–depression association is mediated by adolescent ER.

The Current Study

The assessment design of the current study is multimethod and multisource for measures of parental socialization and adolescent ER, using self-reports as well as observational coding of behaviors during mother–adolescent interactions. We measured early adolescents' ER through the observation of their behaviors in addition to their self-report of ER strategies because observational methods have the potential to capture the dynamics of the ER process and thus contribute unique insights to the understanding of ER compared to methods that focus solely on self-reported ER strategies or affective measures. This study design permits a more comprehensive assessment because each source or method not only provides unique information but also contributes unique error. For example, a sequential observational system permits the detection of microsocial behavioral patterns without the intrusion of perceptual biases inherent in self-report measures, whereas self-report measures provide input from individuals who have substantially greater access than observers to internal regulatory processes and behavior patterns as they occur across time and setting (Sheeber, Davis, Leve, Hops, & Tildesley, 2007). The specific hypotheses for this study were the following:

1. More invalidating (e.g., restrictive, punishing, or dampening) and fewer validating (e.g., encouraging expression) maternal socialization responses to child PA will be associated with higher levels of adolescent depressive symptomatology. Maternal socialization is indicated by maternal self-report and behavioral indices of mothers' responses to their child's PA during mother–adolescent interactions. Given the known associations between parental psychopathology and parenting behaviors (Leinonen, Solantaus, & Punamaki, 2003), maternal depressive and anxiety symptoms were controlled for in all analyses involving maternal socialization.
2. Early adolescents who receive more invalidation and less validation of their PA will display higher levels of emotional dysregulation. Adolescent emotion dysregulation is indicated by greater frequency, duration and reciprocity of negative affective behaviors and lower frequency and duration of positive affective behaviors. It is hypothesized that fewer positive affective behaviors will be reflective of anhedonic tendencies observed in depressed individuals, though this has not been tested empirically to date in adolescents using behavioral observational methodology. Adolescent emotion dysregulation is also indicated by more frequent use of maladaptive ER strategies (as assessed by self-report).
3. Adolescent emotion dysregulation is associated with more depressive symptoms.
4. The relationship between maternal socialization and adolescent depressive symptomatology is mediated by adolescent ER.

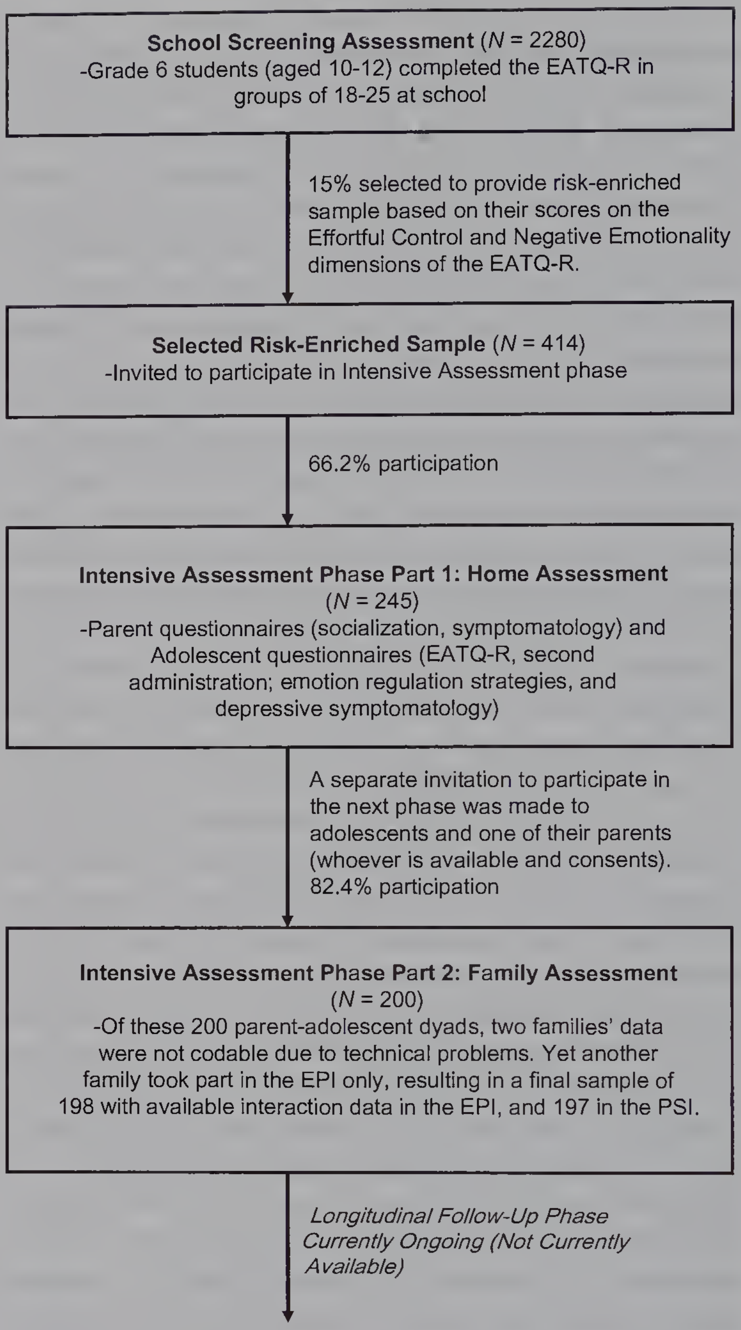
Method

Participants and Selection Procedure

Final-year primary school students across metropolitan Melbourne, Australia, were recruited through their schools to take part in the school screening phase. Through this process, 2,280 students completed Early Adolescent Temperament Questionnaire–Revised (EATQ–R; Ellis & Rothbart, 2001) in the classroom. Given that the main aim of the screening procedure was to provide a risk-enriched sample for the more intensive assessment phase, the EATQ–R scores were used to select a sample that was representative of scores across the full range of the temperament dimensions. Selection was based on scores on Negative Emotionality (NEM) and Effortful Control (EC) dimensions, given their hypothesized role as risk factors for emotional and behavioral disorders. Accordingly, equal numbers of male and female students were selected from each of the following ranges of scores on each dimension: 0–1, 1–2, 2–2.5, and greater than 2.5 *SD* above and below the mean. As outlined in Figure 1, this produced a risk-enriched sample (*N* = 414) for the intensive assessment phase that shows a relatively even distribution across EC and NEM while maintaining the range of temperament scores evident in the larger sample.

The first part of the intensive assessment phase included a battery of questionnaires each for parents and early adolescents to complete. From the selected sample, 245 families consented to participate in this assessment, which occurred a mean of 9.91 months (*SD* = 3.10) after the school screening. At the time of intensive assessment, participating adolescents were 11–13 years old (*M* = 12.46, *SD* = 0.43), and parents' ages ranged from 29 to 57 years (*M* = 42.56, *SD* = 5.19). Although youths of this age range are in their "early adolescence," they are referred to as "adolescents" in this article for ease of expression. Analyses revealed no differences between adolescents who discontinued participation after the school screening phase and those who participated in the intensive phase (Part 1) on NEM, $t(407) = 0.16$, *ns*, or EC, $t(413) = -0.54$, *ns*. The groups also did not differ in gender, $\chi^2(1, N = 415) = 0.34$, *ns*, or socioeconomic classification, $t(405) = -1.00$, *ns*. Socioeconomic classification was based on an index of neighborhood social disadvantage (using participants' residential postal codes), which is calculated as a sum of 13 indicators, such as unemployment, income, criminal offenses, child abuse, and early school leaving (Vinson, 2004).

The 245 families were then invited to take part in the family assessment, and from this, 200 parent–adolescent dyads consented. However, one family completed only the Event-Planning Interaction (EPI), and two sets of data from other participants were not codable. This resulted in 198 sets of behavioral data on the EPI (163 mothers, 80 boys) and 197 (162 mothers, 79 boys) on the Problem-Solving Interaction (PSI). Adolescents who declined to participate in the family assessment did not differ significantly from those who participated in NEM, $t(243) = 0.49$, *ns*; EC, $t(242) = -0.85$, *ns*; gender, $\chi^2(1, N = 245) = 0.15$, *ns*; or mother's country of birth (by continental regions), $\chi^2(7, N = 245) = 12.03$, *ns*. Of the 198 adolescents, 50% were female, and 92% self-identified as "Australian." This sample also did not differ from the original sample (*N* = 2,280) in gender, $\chi^2(1, N = 2,261) = 0.50$, *ns*; or socioeconomic classification, $t(2,241) = 1.55$, *ns*.



School Screening Assessment (*N* = 2280)
-Grade 6 students (aged 10-12) completed the EATQ-R in groups of 18-25 at school

15% selected to provide risk-enriched sample based on their scores on the Effortful Control and Negative Emotionality dimensions of the EATQ-R.

Selected Risk-Enriched Sample (*N* = 414)
-Invited to participate in Intensive Assessment phase

66.2% participation

Intensive Assessment Phase Part 1: Home Assessment
(*N* = 245)
-Parent questionnaires (socialization, symptomatology) and Adolescent questionnaires (EATQ-R, second administration; emotion regulation strategies, and depressive symptomatology)

A separate invitation to participate in the next phase was made to adolescents and one of their parents (whoever is available and consents). 82.4% participation

Intensive Assessment Phase Part 2: Family Assessment
(*N* = 200)
-Of these 200 parent-adolescent dyads, two families' data were not codable due to technical problems. Yet another family took part in the EPI only, resulting in a final sample of 198 with available interaction data in the EPI, and 197 in the PSI.

Longitudinal Follow-Up Phase Currently Ongoing (Not Currently Available)

Figure 1. Sample selection process.
EATQ–R = Early Adolescent Temperament Questionnaire–Revised; EPI = Event-Planning Interaction; PSI = Problem-Solving Interaction.

Given that the sample size of fathers is much smaller than that of mothers, and the exploration of differences in adolescent ER and symptomatology according to parent gender is beyond the scope of this article, all analyses include only the subsample of families where *mothers* participated in the family assessment.

Procedure of Family Assessment Phase

Parents and adolescents participated in two 20-min family interactions, the EPI followed by the PSI (in that fixed order), which were videotaped for coding purposes. Separate cameras videotaped each participant. Order of interactions was fixed because we believed that it would be easier for participants to switch from a positive affective state (as elicited by the EPI) to a conflictual affective state (as elicited by the PSI), given that negative affective states are known to have slower decay rates than positive affective states (Gilboa, 1994). Topics for the EPI and PSI were identified based on participant responses to the Pleasant Events Checklist (PEC) and the Issues Checklist (IC), respectively. The PEC comprises 50 activities that people may enjoy doing, such as "taking a trip or vacation" and "snow skiing," whereas the IC is a list of 44 topics about which adolescents and parents may disagree, such as "[adolescent] lying" and "[adolescent] talking back to parents." Up to 5 activities that both parent and child indicated to be very pleasant in the PEC were chosen for parent–child dyads to plan for during the EPI, and up to 5 topics that were reportedly most conflictual (and recent) on the IC were chosen for the dyads to attempt to resolve during the PSI. As reported in a previous article on this data set (Yap, Allen, O'Shea, et al., 2007), participants displayed more PA in the EPI and more NA in the PSI, indicating a significant task effect. Participants were given $50 cash (parents) and a $30 voucher (adolescents) as reimbursement for their time and travel.

Measures

Parents' Reaction to Children's Positive Emotions Scale

The Parents' Reaction to Children's Positive Emotions Scale (PRCPS; Ladouceur, Reid, & Jacques, 2002) consists of a series of 12 vignettes in which children are likely to experience positive emotions such as joy, pride, excitement, and curiosity. Five of the 12 vignettes have been adapted from the original version (targeted for 4- to 8-year-olds) to be used with older school-age children. The new vignettes were adapted by following the same methodology used to create the original PRCPS for younger children—a committee of experts in developmental psychology and parents (Ladouceur et al., 2002). The PRCPS includes different contexts (with peers or acquaintances, at a birthday party, in a car, etc.), but the parent is always present. Examples of the vignettes used are provided in Appendix A. For each situation, parents indicate on a 7-point scale (from 1 = *very unlikely* to 7 = *very likely*) how likely they would be to react as described in each of the six alternative responses. The PRCPS yields four subscales: Socialization (α = .65), reflecting the degree to which parents explain to their child the reasons why their expressive behavior may be inappropriate given social norms or etiquette; Encouragement (α = .72), indicating the degree to which parents encourage their child to express PA or validate their child's positive emotional states; Reprimand (α = .77), reflecting the degree to which parents react by reprimanding their child for expressing PA; and Discomfort (α = .75), indicating the degree to which parents feel discomfort, embarrassed, or irritated when their child expresses PA. Test–retest reliability of these subscales with preschool children has been reported to range from .60 (Discomfort) to .79 (Reprimand). The convergent/divergent validity of the PRCPS has also been established by significant correlations between its subscales and the scales of other socialization and family environment questionnaires including the Coping with Children's Negative Emotions Scale (Fabes, Eisenberg, & Bernzweig, 1990) and the Family Environment Questionnaire (Moos & Moos, 1981).

For the purpose of data reduction, correlational analyses were conducted to explore whether any of the scales can be collapsed to comprise composite constructs (Appendix A). Analyses revealed that Encouragement was negatively correlated with the other three scales (*r*s = −.27 to −.49); therefore, it is termed *validating socialization,* reflecting nonrestrictive responses to child positive affective displays. The three remaining scales were positively correlated with one another (*r*s = .34 to .66), reflecting parental discomfort and restrictive, dampening responses to child PA. The mean of these scales was calculated and used as an overall index of invalidating socialization of PA, which has a high internal consistency of α = .86.

Child Affect Questionnaire–Child Strategies

The Child Affect Questionnaire–Child Strategies (Garber et al., 1995) was used to assess the frequency with which adolescents engaged in various responses for regulating NA in an interpersonal scenario. Adolescents were instructed to "imagine that you had just had a big fight with a good friend" and to indicate how often they would use each of the strategies after the event, from 1 (*none of the time*), 2 (*a little of the time*), 3 (*about half of the time*), or 4 (*a lot of the time*). The strategies were grouped into six categories based on the existing literature concerning categories of affect regulation and coping strategies (e.g., Band & Weisz, 1988). However, only the Negative Responses scale was used in this study as a measure of "maladaptive ER" strategies, as this scale is conceptually and empirically indicative of poor ER (e.g., N. B. Allen, Sheeber, Leve, Davis, & Golston, 2003). The Negative Responses scale consisted of nine strategies, including "talk back to parents" and "continue to argue with the friend," and had a high Cronbach's alpha of .80. These strategies reflect deficits in managing the NA experienced in the given conflict scenario, resulting in inappropriate venting and dysregulated expressions of NA. Although almost 18% of adolescents reported using these strategies "none of the time," 71% of them reported using at least some of these strategies "a little of the time" (*M* = 13.47, *SD* = 4.34).

Center for Epidemiological Studies Depression Scale – Revised

Maternal and adolescent depressive symptomatology was measured using self-reports on the Center for Epidemiological Studies Depression Scale – Revised (CES-D – R; Radloff, 1977). The CES-D – R has been found to be valid and reliable for adults and adolescents, with Cronbach's alpha of .85 reported in both a general population sample of adults (Radloff, 1977) and a nonclinical sample of junior high school students (Radloff, 1991). Most of this adolescent sample reported few depressive symptoms (*M* = 11.44, *SD* = 9.38), with 59% having total scores of 0 – 10. Approximately 11% (14 male and 11 female) had total scores of 24 and above, suggesting that they have a relatively high likelihood of meeting criteria for major depressive disorder (Roberts, Lewinsohn, & Seeley, 1991). Many mothers also reported few depressive symptoms (*M* = 10.37, *SD* = 10.09), with 62% having total scores ≤ 10, although 19% reported total scores exceeding the adult clinical cutoff of 16. Cronbach's alphas were .89 and .92 for adolescents and parents, respectively, in this study.

Beck's Anxiety Inventory

A state measure of maternal anxiety symptoms was obtained by maternal self-report on the Beck's Anxiety Inventory (BAI; Beck & Steer, 1990). The BAI has demonstrated good psychometric properties in a nonclinical college sample, showing a high internal consistency with Cronbach's alpha of .90, which is comparable to that in this sample (α = .92). Using the guidelines of Beck and Steer (1990) for score ranges, 70% of mothers reported a "normal" anxiety level, 23% mild to moderate, 3% moderate to severe, and 4% severe anxiety.

Family Observation Measure

The affect and verbal content of the videotaped interactions were coded in real time using the Living in Family Environments (LIFE; Hops, Davis, & Longoria, 1995) coding system. The LIFE is an event-based coding system in which new codes are entered each time the affect or verbal content of the interactants changes. The LIFE consists of 10 affect codes and 27 verbal content codes. Three composite variables, derived from the individual affect and content codes, were used in the current study (see Appendix B for the full list of codes and composition of constructs). *Aversive* behavior includes all codes with contemptuous, angry, and belligerent affect, as well as disapproving, threatening, or argumentative statements with neutral affect. *Dysphoric* behavior consists of all codes with dysphoric, anxious, or whining affect, as well as complaints and self-derogatory comments with neutral affect. A new Positive Interpersonal behavioral construct was developed for this study, consisting of statements made with happy, pleasant, and caring affect, as well as approving or affirming statements, and statements that serve to maintain the conversation. The validity of the former two constructs has been established in earlier studies examining parent – child interactions (e.g., Hops, Sherman, & Biglan, 1990; Sheeber et al., 2000). More detailed information about the development and psychometric characteristics of LIFE is presented in Hops et al. (1995).

All video recordings were coded by extensively trained observers who are blind to participant characteristics (e.g., temperament or symptomatology levels). Approximately 20% of the interactions were coded by a second observer to provide an estimate of observer agreement. Kappa scores in this study, which were in the good to excellent range (Fleiss, 1981), averaged .66 and .78 for content and affect codes, respectively, and kappa coefficients were .77, .68, and .89 for the Aversive, Dysphoric, and Positive Interpersonal composite codes, respectively. Random pairs of observers were assigned to the interactions to minimize drift between any two observers and to ensure that all observers met minimal criteria for acceptable observations. The base rate for coding an affect or content code was 1, which means that when it fails to occur at least once, it arises as a missing value for that code for that particular participant. This results in variations in *N* for different constructs in different tasks.

On the basis of data obtained from the LIFE coding of family interactions, behavioral indices of maternal socialization and adolescent ER were derived. These include the following.

Rate per minute (rpm). The rpm variable indicates the number of times behaviors associated with a composite construct occurs per minute of interaction between participants. Separate variables were constructed to assess the average frequency of Aversive, Dysphoric, and Positive Interpersonal displays. More frequent aversive and dysphoric behaviors and less frequent positive interpersonal behaviors are taken as indicative of emotion dysregulation. The rpm variable has been more commonly used as a measure of positivity, negativity, and problem-solving capacity in family and parent–child interactions (e.g., Forgatch, 1989; Jacob & Johnson, 2001) and broadly reflects the emotional climate in these relationships. In this study, rpm is used as a broad measure of ER in that it indexes how easily one is drawn into a particular affective state (emotion generation/reactivity), as well as one's capacity in resisting the particular affective state (ER) and hence not produce it as often.

Sequential z score. The *z* scores for sequential relations indicate the extent to which a specified sequence of behavior occurs more or less often than would be expected as a function of the base rate of each behavior. The *z* scores represent whether a particular antecedent behavior is effective in eliciting/increasing the likelihood of (positive *z* score) or suppressing (negative *z* score) a particular consequent behavior (see Hops et al., 1995, for a more detailed description).

Specific behavioral sequences are of interest in this study as indicators of adolescent ER. In particular, greater reciprocity of aversive and dysphoric behaviors (e.g., positive *z* scores of adolescent aversive responses to maternal aversive behavior) is taken as indicative of emotion dysregulation. In creating these variables, antecedent behaviors are regarded as provocative affective stimuli. The extent to which the respondent reciprocates NA is therefore considered to be an index of the respondent's ability to regulate (e.g., diffuse) the affect generated by negative antecedent behaviors. This methodology is consistent with the manner in which difficulties in ER have been assessed in the context of both conflictual marital exchanges (e.g., Lindahl & Markman, 1990) and problem-solving interactions between adolescents (depressed and nondepressed groups) and their mothers (Sheeber et al., 2000).

Invalidating maternal socialization is indexed by a greater likelihood (i.e., more positive *z* scores) of "punishing" (aversive) and "dampening" (dysphoric) maternal responses following adolescent positive interpersonal behavior. Given that there has been minimal research on socialization of PA in adolescents, the punishing and dampening parental response variables were created as a conceptual parallel to maladaptive socialization of NA described in the theoretical and empirical literature (e.g., Eisenberg, Cumberland, & Spinrad, 1998; Eisenberg et al., 1996), which is characterized by punitive, dismissive parental responses to child affect.

Duration per episode (dpe). Dpe represents the average length of time that specific affective behaviors are maintained each time they are displayed during the course of each mother–adolescent interaction task. It is developed to capture the person's ability to shift out of negative affective states as well as to maintain positive states. Separate variables were constructed to assess the average duration of aversive, dysphoric, and positive interpersonal displays. A longer duration of aversive and dysphoric behaviors and a shorter duration of positive interpersonal behaviors are taken as indicative of emotion dysregulation.

To make analyses more methodologically stringent when using maternal behavior variables (as indices of maternal socialization) to predict adolescent behavior variables (as indices of adolescent ER), mother and child behavioral indices are derived from separate interaction tasks. In particular, given our interest in maternal socialization in the normal family environment, as a reflection of the type of parenting the child is usually exposed to, we examine maternal socialization during the nonconflictual, positive affective interaction context (EPI). Moreover, the EPI is more likely to evoke more PA from the adolescent, and it provides more opportunity to examine mothers' socialization responses to such. On the other hand, given our interest in adolescents' ability to regulate their emotions specifically in conflictual contexts (parent–child and/or peer), we examine adolescent ER during the PSI, which presumably poses challenges to the adolescents' ER capacity. In addition, adolescent ER in interpersonal conflictual contexts is also of interest because difficulties in interpersonal relationships have been associated with adolescent depression (Eisenberg et al., 1996, 1998). However, when predictor and outcome variables are derived from different methods (e.g., parent-report socialization or adolescent-report ER strategies with behavioral variables), behaviors in both tasks will be examined.

Results

Given the ample evidence from the developmental literature indicating adolescent gender differences in ER and depressive symptomatology (e.g., Hankin et al., 1998; Thayer, Rossy, Ruiz-Padial, & Johnsen, 2003) and in sensitivity to parental psychopathology and disturbances in parent–child interactions (e.g., Chaplin, Cole, & Zahn-Waxler, 2005; Klimes-Dougan et al., 2007), any moderating effects of gender on the associations between maternal socialization and adolescent outcomes are examined when testing the hypotheses of this study. Because few prior studies have investigated socialization and ER behaviors during parent–adolescent interaction tasks designed to elicit PA, behaviors in the EPI and PSI are examined separately to see if different forms of socialization and emotion dysregulation are evident in different parent–adolescent interaction contexts. All hypotheses were tested using hierarchical regressions, with statistical significance set at $\alpha = .05$. To test the first two hypotheses, adolescent gender was entered in Step 1, maternal depressive and anxiety symptoms in Step 2, maternal socialization in Step 3, and Socialization × Gender interaction in Step 4. To test Hypothesis 3, analyses involved Steps 1, 3, and 4 given above and the predictor variable is adolescent ER. When testing the effect of adolescent ER (mediator) on symptomatology, maternal socialization (predictor) was controlled for as well.

Following the recommendation of Baron and Kenny (1986), the mediation model requires three criterion relationships, namely, (1) predictor (maternal socialization) variable exerts significant effects on mediator (adolescent ER), (2) outcome variable (depressive symptoms), and (3) mediator exerts significant effects on outcome variable. As such, where any of these three relationships was not significant in any model, mediation effects will not be tested. Although some scholars have more recently argued that Criterion (2) should not be required for mediation (MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002), we chose to set these criteria because the conceptual model being tested in this study (Yap, Allen, & Sheeber, 2007) would only be meaningful if these relationships are significant. The significance of any mediation effect will be tested using the Sobel test (Aroian version; Preacher & Leonardelli, 2001), and indirect or mediated effects are calculated following the recommendations of MacKinnon and Warsi (1995). This calculation involves a two-stage regression procedure: First, the mediator (adolescent ER) is regressed on the predictor (maternal socialization). Second, adolescent symptomatology is regressed on both predictor and mediator variables. The product of the regression coefficient of the predictor from the first regression with the coefficient of the mediator in the second regression yields the mediated effect of maternal socialization on adolescent symptomatology.

Adolescent gender had no significant main effects on their ER or depressive symptoms. Maternal symptomatology also had no main effect on adolescent symptomatology, but maternal depressive symptoms were positively associated with adolescent reciprocity of aversive behaviors during the PSI, $\beta = .22$, $t(155) = 200$, $p < .05$, $\Delta R^2 = .04$. All analyses are summarized in Tables 1 and 2.

Predicting Adolescent Depressive Symptomatology With Maternal Socialization

Regression analyses were first conducted to test the associations that adolescent depressive symptomatology has with mothers' socialization responses to adolescent PA. As predicted, maternal dampening of adolescent positive behaviors (i.e., a positive conditional probability for maternal dysphoric behavior given adolescent PA) during both interaction tasks was associated with more symptomatology. Adolescent gender moderated the association between maternal dampening in the EPI and adolescent symptoms, and follow-up analyses revealed that this association was significant for females but not for males, $\beta = .37$, $t(72) = 3.31$, $p < .005$, $\Delta R^2 = .13$. Gender moderation was also found for the association between maternal self-report of invalidating socialization and adolescent symptoms, with the association being significant for females only, $\beta = .32$, $t(75) = 2.97$, $p < .005$, $\Delta R^2 = .10$. Maternal punishing responses during mother–adolescent interactions and maternal report of validating socialization of PA were unrelated to adolescent depressive symptomatology.

Predicting Adolescent ER With Maternal Socialization

Consistent with expectations, adolescents whose mothers report being prohibitive of their PA displays (invalidating socialization) were more likely to reciprocate their mothers' aversive behaviors. Validating socialization was not directly associated with adolescent ER behaviors, but adolescent gender moderated some of these associations, which were significant for males but not for females. Male adolescents whose mothers report being more likely to encourage their positive affective displays demonstrated better ER during the PSI. These boys behaved aversively less often, $\beta = -.30$, $t(72) = -2.65$, $p < .01$, $\Delta R^2 = .08$; maintained their positive behaviors for a longer

Table 1

Summary of Hierarchical Regressions Predicting Adolescent Depressive Symptomatology With Maternal Socialization and Adolescent Emotion Regulation Variables

	N	B	$SE\ B$	β	ΔR^2
Maternal socialization responses to Ad positive affect					
Mo aversive given Ad positive (EPI)	148	0.36	0.63	.05	.00
Mo dysphoric given Ad positive (EPI)	150	2.49	0.85	.24†	.06
Mo dysphoric Given Ad positive × Gender		−3.87	1.71	−.25*	.03
Mo aversive given Ad positive (PSI)	151	0.57	0.49	.10	.00
Mo dysphoric given Ad positive (PSI)	149	2.31	0.81	.23†	.05
Mo self-report validating responses (PRCPS)	155	−0.07	0.07	.08	.00
Mo self-report invalidating responses (PRCPS)	155	0.15	0.09	.14	.02
Mo Self-Report Invalidating Responses × Gender		−0.50	0.18	−.31**	.05
Ad ER behaviors (PSI)					
Aversive reciprocity	151	0.51	0.24	.17**	.03
Aversive Reciprocity × Gender		−1.08	0.47	−.27	.03
Dysphoric reciprocity	149	0.49	0.56	.07	.01
Ad ER strategies (self-report)					
Maladaptive ER strategies	134	1.23	0.15	.57††	.33

Notes. These values are based on models controlling for Ad gender (Step 1) and maternal symptomatology (Step 2; no significant main effects and thus not shown here), the above-mentioned variables in Step 3, and the interaction between gender and these variables in the final step (only significant interactions are shown here). The analyses using the frequency and duration of Ad behaviors in both tasks and Ad reciprocity in the EPI as predictors are nonsignificant and hence are not shown here (due to space constraints). A total of 21 regression analyses were conducted. ΔR^2 refers to change in R^2 values from Steps 2 to 3, except where interaction effects are shown, where the change is from Steps 3 to 4. Ad = adolescent; EPI = Event-Planning Interaction; ER = emotion regulation; Mo = mother; Positive = positive interpersonal; PRCPS = Parents' Reaction to Children's Positive Emotions Scale; PSI = Problem-Solving Interaction.
$^{*}p < .05$. $^{**}p < .01$. $^{\dagger}p < .005$. $^{\dagger\dagger}p < .001$.

duration, $\beta = .30$, $t(72) = 2.65$, $p < .01$, $\Delta R^2 = .09$; and were less likely to reciprocate their mothers' aversive behaviors, $\beta = -.30$, $t(72) = -2.82$, $p < .01$, $\Delta R^2 = .09$.

As predicted, adolescents whose mothers were more likely to punish their positive behaviors in the EPI displayed aversive behaviors more often in the PSI. Some gender effects were also found among these adolescents: Males maintained their aversive behaviors for a longer duration, $\beta = .33$, $t(68) = 2.94$, $p < .005$, $\Delta R^2 = .11$, and females were more likely to reciprocate their mothers' dysphoric behaviors, $\beta = .30$, $t(77) = 2.74$, $p < .01$, $\Delta R^2 = .09$. Maternal dampening of adolescent positive behaviors was unrelated to adolescent ER behaviors, and adolescent PA regulation was unrelated to all measures of invalidating maternal socialization.

Adolescents whose mothers reported more invalidating socialization reported using maladaptive ER strategies more frequently. This association was moderated by adolescent gender, and follow-up analyses revealed that it was significant for females only, $\beta = .44$, $t(65) = 4.00$, $p < .001$, $\Delta R^2 = .19$. Adolescents whose mothers were more likely to dampen their positive behaviors during the PSI also reported using maladaptive ER strategies more often. Maternal punishing of adolescent positive behaviors was not associated with adolescent ER strategies.

Predicting Adolescent Depressive Symptomatology With Adolescent Emotion Dysregulation

Contrary to expectations, the frequency (rpm) and duration (dpe) of adolescent Aversive, Dysphoric, and Positive Interpersonal behaviors in both the EPI and the PSI were all not significantly associated with depressive symptomatology. However, adolescents with more depressive symptoms are more likely to reciprocate their mothers' aversive behavior in the PSI and their mothers' dysphoric behavior in the EPI. Moreover, adolescent gender moderated the former association, and follow-up analyses revealed that it was significant for females only, $\beta = .33$, $t(74) = 3.01$, $p < .005$, $\Delta R^2 = .11$. More depressed adolescents also reported using maladaptive ER strategies more frequently when they experience NA.

Testing for Mediation

Following the recommendations of Baron and Kenny (1986), the only potential mediator to be tested here is maladaptive ER strategies, in the associations that maternal dampening behaviors in the PSI, and for females only, maternal report of invalidating socialization, have with adolescent depressive symptomatology because these are the only sets of associations that fulfill all three criteria.

Table 2

Summary of Hierarchical Regressions Predicting Adolescent Emotion Regulation With Maternal Socialization Variables

DV: Ad ER	*N*	*B*	*SE B*	β	ΔR^2
IV: Mo aversive given Ad positive (EPI)					
Aversive frequency (rpm; PSI)	154	0.10	0.04	.22**	.05
Positive frequency (rpm; PSI)	154	0.05	0.05	.10	.01
Aversive duration (dpe; PSI)	154	0.07	0.68	.01	.00
Positive duration (dpe; PSI)	154	−0.27	0.20	−.11	.01
Aversive reciprocity (PSI)	154	0.40	0.20	.16	.02
Dysphoric reciprocity (PSI)	152	0.17	0.09	.15	.02
Dysphoric Reciprocity × Gender		0.37	0.18	.23*	.03
Maladaptive ER strategies	133	0.05	0.32	.01	.00
IV: Mo aversive given Ad positive (PSI)					
Maladaptive ER strategies	136	0.32	0.25	.11	.01
IV: Mo dysphoric given Ad positive (PSI)					
Maladaptive ER strategies	134	1.28	0.38	.28†	.08
IV: Mo self-report validating responses (PRCPS)					
Aversive frequency (rpm; PSI)	158	−0.00	0.00	−.06	.00
Aversive Frequency × Gender		−0.03	0.01	−.33**	.05
Positive frequency (rpm; PSI)	158	−0.00	0.01	−.02	.00
Aversive duration (dpe; PSI)	158	−0.03	0.08	−.03	.00
Positive duration (dpe; PSI)	158	0.04	0.02	.12	.02
Positive duration × Gender		0.12	0.05	.29*	.04
Aversive reciprocity (PSI)	158	−0.03	0.02	−.11	.01
Aversive Reciprocity × Gender		−0.12	0.05	−.30*	.04
Maladaptive ER strategies	139	−0.00	0.04	−.00	.00
IV: Mo self-report invalidating responses (PRCPS)					
Aversive frequency (rpm; PSI)	158	0.01	0.01	.10	.01
Positive frequency (rpm; PSI)	158	0.00	0.01	.00	.00
Aversive duration (dpe; PSI)	158	0.07	0.10	.06	.00
Positive duration (dpe; PSI)	158	−0.05	0.03	−.14	.02
Aversive reciprocity (PSI)	158	0.07	0.03	.19*	.03
Maladaptive ER strategies	139	0.17	0.04	.31††	.09
Maladaptive ER Strategies × Gender		−0.22	0.09	−.27*	.04

Notes. These values are based on models controlling for Ad gender (Step 1; no significant main effects and thus not shown here) and maternal symptomatology (Step 2; only one significant main effect, described in text), the above-mentioned variables in Step 3, and the interaction between gender and these variables in Step 4 (only significant interactions are shown here). ΔR^2 refers to change in R^2 values from Steps 2–3, except where interaction effects are shown, where the change is from Steps 3 to 4. Because most analyses involving Ad dysphoric behaviors and maternal socialization in the EPI yielded nonsignificant findings, only those with significant results are shown. A total of 38 regression analyses were conducted. Ad = adolescent; dpe = duration per episode; DV = dependent variable; EPI = Event-Planning Interaction; ER = emotion regulation; IV = independent variable; Mo = mother; Positive = positive interpersonal; PRCPS = Parents' Reaction to Children's Positive Emotions Scale; PSI = Problem-Solving Interaction.
*$p < .05$. **$p < .01$. †$p < .005$. ††$p < .001$.

A series of hierarchical regressions revealed that maladaptive ER strategies accounted for unique variance in symptomatology even after each of the two above-mentioned predictors was entered into the regression model; therefore, the Sobel test was conducted to test for the significance of the mediation effects. As shown in Figure 2, the effects of both socialization variables on adolescent symptomatology became nonsignificant after accounting for the effects of adolescent ER strategies, indicating a significant mediation effect in both cases.

Discussion

On the whole, all four hypotheses were at least partially supported. Maternal dampening of adolescent positive affective behaviors, and in females only, maternal report of invalidating PA socialization, were associated with more adolescent depressive symptoms. Adolescents whose mothers were more likely to punish and dampen their positive affective behaviors during mother–adolescent interactions and to report using invalidating socialization of their PA expression demonstrated more emotion dysregulation.

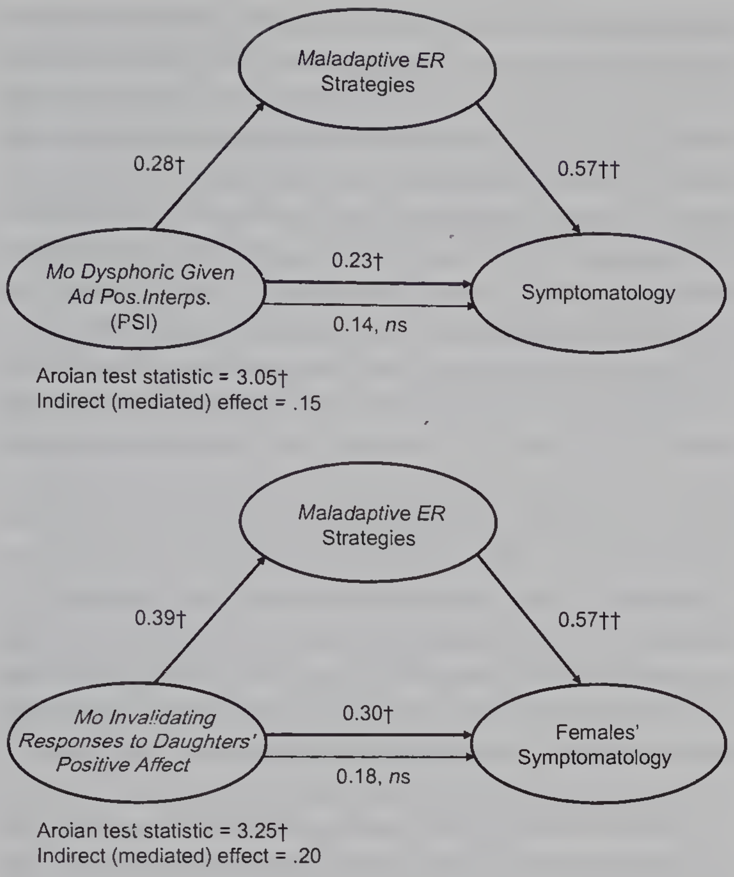

Maladaptive ER Strategies
0.28†
0.57††
Mo Dysphoric Given Ad Pos.Interps. (PSI)
0.23†
0.14, ns
Symptomatology
Aroian test statistic = 3.05†
Indirect (mediated) effect = .15
Maladaptive ER Strategies
0.39†
0.57††
Mo Invalidating Responses to Daughters' Positive Affect
0.30†
0.18, ns
Females' Symptomatology
Aroian test statistic = 3.25†
Indirect (mediated) effect = .20

Figure 2. Diagram showing significant mediation of maladaptive ER strategies in the relationship between maternal dampening of Ad positive behavior during the PSI (top panel) and depressive symptomatology, and maternal self-report of Invalidating Socialization and symptomatology, females only (bottom panel).
Note. Values shown are standardized regression coefficient (β) values. Ad = adolescent; ER = emotion regulation; Mo = mother; Pos. Interps = positive interpersonal; PSI = Problem-Solving Interaction.
$^{\dagger}p < .005$. $^{\dagger\dagger}p < .001$.

Moreover, adolescents with more dysregulated behaviors and who use maladaptive strategies more frequently to manage their NA reported more depressive symptoms. Adolescent ER strategies, but not behaviors, significantly mediated the association between maternal invalidation of PA and adolescent depressive symptomatology. Interestingly, validating socialization of PA was related to adolescent ER in males only and was not associated with adolescent depressive symptoms.

Maternal Socialization of PA and Adolescent Depressive Symptomatology

Findings from multimethod, multiinformant measures of maternal socialization partially supported the first hypothesis that invalidating maternal PA socialization predicts more adolescent self-reported depressive symptomatology. However, validating socialization was found to be unrelated to symptomatology. The finding that maternal dampening responses to adolescent positive affective behaviors, as well as maternal report of invalidating PA socialization, are associated with elevated depressive symptomatology in early adolescents extends and replicates findings of recent, unpublished studies (Katz, 2002; Sheeber & Davis, 2002, as described and cited in Sheeber, Allen, Davis, & Katz, 2004). In particular, it is noteworthy that negative socialization of PA has implications for adolescent mental health.

Maternal dampening of adolescent positive behaviors seems to be an important form of maladaptive socialization that warrants further study because most studies of socialization have focused on punishing forms of parental socialization responses. The association between maternal dysphoric responses to their children's positive behaviors resembles those of depressed mothers (e.g., Hops et al., 1990; Jacob & Johnson, 1997); therefore, it may be argued that these mothers' tendency to dampen their adolescents' positive affective displays is related to the mothers' heightened symptomatology. However, our findings demonstrate that maternal socialization has effects on adolescent ER and depressive symptoms over and above any main effects of maternal symptomatology, which were minimal. Further post hoc analyses conducted to examine whether such maternal socialization responses are related to mothers' ER also revealed no significant associations. Therefore, it seems that mothers' likelihood to respond in a dysphoric manner to their adolescents' positive interpersonal behaviors has less to do with the mothers' own emotional functioning than it does with the interpersonal interaction context and the adolescents' emotional functioning.

Maternal invalidating socialization of adolescent PA (including maternal discomfort toward and tendency to use reprimanding or socializing emotion management strategies to decrease their child's expression of PA) was significantly associated with adolescent symptomatology, but only in females. Given that the effect of PA socialization on adolescent depressive symptomatology has not been studied before in past research, it is unclear why no significant effect was found in this study for males, and why maternal punishing behaviors were unrelated to adolescent symptoms as well. It is possible that such socialization behaviors contribute to adolescent depressive symptomatology indirectly via their impact on adolescent ER, namely, increasing the likelihood that these adolescents would reciprocate their mothers' aversive and dysphoric behaviors and use maladaptive ER strategies more often. Invalidating PA Socialization may also have an impact on boys that was not captured by the outcomes examined in this study, for example, on externalizing problems. Likewise, maternal punishing socialization may contribute directly or indirectly to externalizing problems. Nonetheless, in the absence of the replication of these findings, it is difficult to draw firm conclusions.

Maternal Socialization of PA and Adolescent ER

On the whole, as predicted, more maternal invalidation and less validation of adolescent PA, as measured by multimethod, multisource methods, was found to be associated with poorer ER in the adolescent. The associations underscore the impor tance of investigating the previously neglected research area of the socialization of *positive* emotions for a better understanding of the effects of socialization on adolescent ER. Maternal responses that are restrictive and unaccepting of positive affective displays in their children (mother-report invalidating socialization) are associated with a greater likelihood of adolescents reciprocating their mothers' aversive behaviors in the PSI and more frequent use of ineffective ER strategies to manage NA elicited from a peer conflict situation. These findings are consistent with those of Denham et al. (1997), whereby parents' use of guiding and socializing language (similar to PRCPS Socialization and Reprimand scales) was negatively associated with preschool children's emotional and social competence, suggesting that parents may discourage affective displays of children who most actively need such tutelage (i.e., who are seemingly less able to self-regulate). Nonetheless, it is important to note that these cross-sectional associations may differ from longitudinal associations; hence, future research is warranted. Being more easily pulled into aversive exchanges with their mothers is a marker of poor ER for adolescents, given that it suggests an inability to downregulate their aversive affective reaction when provoked by maternal aversive behavior, in this case in a conflict resolution interaction with their mother. Indeed, these adolescents may have a tendency to use maladaptive strategies like venting and dysregulated expression of NA in such conflict situations, whether it is with their mother or their peers.

It is noteworthy that validating PA socialization was associated with better ER in males but not in females. Post hoc analyses revealed that this gender difference was not due to differences in mothers' likelihood to encourage PA expression in girls versus boys. Given the lack of past research into the role of PA validation in adolescent ER, or gender differences in this association, more research is required before further inferences can be drawn. Nonetheless, it is interesting that mothers' validation of their sons' PA is associated with male adolescents' regulation of both positive and negative affective behaviors during a conflict resolution interaction with their mothers. This is broadly consistent with evidence that parental sharing of PA with young children is associated with children's social competence (Denham, 1997). If replicated, these findings have important implications for parenting interventions, suggesting that teaching parents to encourage their sons' PA expression may contribute to their sons' development of better ER.

This study used two behavioral measures of invalidating PA socialization that parallel the mother-report invalidating socialization measure, namely, maternal punishing (aversive) and dampening (dysphoric) responses to adolescent positive interpersonal behaviors. On the whole, the associations that these behavioral measures of socialization have with multimethod, multiinformant indices of adolescent ER are consistent with those involving questionnaire measures of socialization—maternal invalidation of adolescent positive affective displays engender risk for adolescent emotion dysregulation.

Adolescents whose mothers were more likely to respond aversively toward them when they displayed positive interpersonal behavior during the EPI (a context where such behaviors are normative and adaptive, thus making maternal punishing responses somewhat unjustified) displayed aversive behaviors more frequently during the PSI. Amongst these adolescents, males also maintained their aversive behaviors for a longer duration, whereas females were more likely to reciprocate their mothers' dysphoric behaviors. Maternal dampening of adolescent positive interpersonal behaviors was not related to adolescent ER *behaviors* but was associated with adolescents using maladaptive strategies more frequently to manage their NA.

In sum, findings from multimethod, multiinformant measures of maternal socialization of PA and adolescent ER paint a coherent picture consistent with past research involving NA and younger children (e.g., Denham et al., 1997) and with theoretical expectations involving PA. Maternal invalidation of adolescent PA comprises dysfunctional socialization in that they engender risk for adolescent dysregulation of both PA and NA. In particular, this study is among the first to investigate and demonstrate the effects of parental socialization of PA on adolescent ER.

Adolescent ER and Depressive Symptomatology

Adolescents with more depressive symptoms were more likely to reciprocate their mothers' aversive behaviors during the PSI and dysphoric behaviors during the EPI. Such individual differences in aversive and dysphoric reciprocity during early adolescence may be indicators of risk for depressive disorders later in adolescence, arguably because such exchanges underlie poor ER in these early adolescents (Sheeber et al., 2000). In addition, adolescents who use maladaptive, dysregulated expressions and avoidance behaviors more frequently as strategies for regulating their negative emotions presented with elevated depressive symptoms. This latter finding is consistent with research on the ER strategies of adolescents and young adults (e.g., Gross & John, 2003; Silk et al., 2003).

Adolescent ER as a Mediator

Consistent with hypothesis, maternal dampening of adolescent PA predicted adolescent symptomatology through its effects on adolescent ER strategies but not on ER behaviors. This may be because self-report of strategies better reflects trait-like aspects of adolescent ER and is better placed to capture those aspects of adolescent functioning that are consistent over time and are therefore able to mediate the influence of parenting on symptoms. In females only, adolescent ER strategies also mediated the association between mother-report invalidating socialization and depressive symptoms. It is notable, moreover, that the effect of socialization on symptoms dropped to nonsignificance after adding the mediator to the model, suggesting that maladaptive ER strategies are a potent mediator in the association between maternal PA socialization and adolescent depressive symptomatology.

In particular, mothers who respond in an anxious or dysphoric manner (dampening) to their adolescent's positive interpersonal behaviors during the EPI may confer greater risk for depressive symptomatology in these adolescents because they use ineffective strategies more frequently to cope with their NA. The same applies for female adolescents whose mothers express discomfort, socialize, and/or reprimand them for expressing PA (invalidating socialization). In essence, such unjustified discouragement and dampening of their positive interpersonal behaviors may leave these adolescents frustrated and angry yet unable to successfully downregulate these emotions; hence, they experience a buildup of such emotions that can become pervasive and in turn lead to the elevation of depressive symptoms. Alternatively, when frustrated and angry, these adolescents may be less able to upregulate their PA to offset their NA or to diffuse the tension in the parent – adolescent relationship, which in turn contributes to their risk for developing more severe depressive problems.

Gender Differences

Some interesting gender differences emerged in this study, particularly in the associations between maternal PA socialization and adolescent ER and depressive symptoms. Overall, it appears that female adolescents' risk for depression is particularly sensitive to the effects of maternal invalidation of their PA. Researchers have suggested that this may be due to the intensity of the mother – daughter relationship in particular (Youniss & Ketterlinus, 1987) and due to increasing societal demands on them to maintain cordial interpersonal relationships (Davies & Windle, 1997). Notably, these findings suggest that this vulnerability may be evident from early adolescence. An implication of these findings is that parents should be cautious when responding to their daughters' PA even within apparent socialization norms, as such restrictive responses to girls' PA may have adverse effects on their emotional and psychological functioning. On the other hand, boys seem to benefit from maternal validation of their PA expressions, particularly in terms of their ability to regulate both their PA and NA. Hence, although boys are adversely affected by maternal invalidation, these findings suggest that maternal validation may serve as a protective factor for male early adolescents.

Strengths and Limitations

The current study has a number of notable strengths. It is based on a large population sample of adolescents, from which a smaller risk-enriched subsample was selected to represent a wide variance in temperament. Moreover, this study used a multimethod and multisource measure of ER, including a behavioral measure of adolescent ER and a self-report measure of ER strategies. This allowed the examination of multiple dimensions of emotions: valence (PA and NA), frequency, duration, and reciprocity, as recently recommended by researchers in the field (Gross & John, 2003; Larsen, 2000). Furthermore, this study addressed an important gap in the literature by examining the direct effects of maternal socialization of PA on ER and depressive symptoms in early adolescents. It also derived indices of maternal socialization and adolescent ER behaviors from separate interaction tasks and used multimethod and multisource measures of maternal socialization.

There were also limitations to this study. Only mothers were included in this study; hence, the potential differences in the effects of parental socialization due to parent gender could not be examined. The associations reported here are cross-sectional;

thus, findings from the longitudinal follow-up are required to clarify the direction of associations and to examine the possibility that ER mediates the association between parental socialization of PA and depression later in adolescence. In particular, longitudinal data are required to make more rigorous inferences about the causal relations implied in this mediational model (D. A. Cole & Maxwell, 2003). In addition, the risk-enriched sampling design of this study may have inflated effect sizes relative to a normal community sample.

Conclusions

Maternal socialization of adolescent PA has important implications for adolescents' ER strategies and their affective behaviors during mother–adolescent interactions, as well as their concurrent depressive symptoms. Adolescents with more depressive symptoms tend to use maladaptive ER strategies more frequently and display more emotion dysregulation during mother–adolescent interactions. Mothers' invalidation of adolescent PA and adolescents' emotionally dysregulated behaviors are likely to create and contribute to a family emotional climate that limits the adaptive development of ER skills in adolescents and heightens their risk for more severe depressive problems. These findings, albeit preliminary, highlight the importance of focusing on PA when working with parents of young adolescents at risk for depression. For example, parents should be made aware of the impact that preventing their children from experiencing or expressing PA can have on their children's development of ER and their risk for affective problems. It remains to be seen if adolescent ER behaviors would mediate the longitudinal association between their mothers' socialization behaviors and their diagnostic status, namely, of major depression, later in adolescence.

References

Allen, J. P., Hauser, S. T., Eickholt, C., Bell, K. L., & O'Connor, T. G. (1994). Autonomy and relatedness in family interactions as predictors of expressions of negative adolescent affect. *Journal of Research on Adolescence*, *4*, 535–552.

Allen, N. B., Sheeber, L., Leve, C., Davis, B., & Golston, S. (2003). Autonomic cardiac control and family interactions: Adolescent affect and affect regulation [Abstract]. *Psychophysiology*, *40*, S22.

Astington, J. W. (1996). What is theoretical about the child's theory of mind? A Vygotskian view of its development. In P. Carruthers & P. Smith (Eds.), *Theories of theories of mind* (pp. 184–199). Cambridge: Cambridge University Press.

Band, E. B., & Weisz, R. (1988). How to feel better when it feels bad: Children's perspectives with everyday stress. *Developmental Psychology*, *24*, 247–253.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality & Social Psychology*, *51*, 1173–1182.

Barrera, M., & Garrison-Jones, C. (1992). Family and peer social support as specific correlates of adolescent depressive symptoms. *Journal of Abnormal Child Psychology*, *20*, 1–16.

Beck, A. T., & Steer, R. A. (1990). *Beck Anxiety Inventory manual*. San Antonio, TX: The Psychological Corporation Harcourt Brace Jovanovich.

Calkins, S. D., & Bell, K. L. (1999). Developmental transitions as windows to parental socialization of emotion. *Psychological Inquiry*, *10*, 368–372.

Campos, J. J., Frankel, C. B., & Camras, L. (2004). On the nature of emotion regulation. *Child Development*, *75*, 377–394.

Cassidy, J. (1994). Emotion regulation: Influences of attachment relationships. *Monographs of the Society for Research in Child Development*, *59*, 228–283.

Chaplin, T. M., Cole, P. M., & Zahn-Waxler, C. (2005). Parental socialization of emotion expression: Gender differences and relations to child adjustment. *Emotion*, *5*(1), 80–88.

Clark, L. A., & Watson, D. (1991). Tripartite model of anxiety and depression: Psychometric evidence and taxonomic implications. *Journal of Abnormal Psychology*, *100*, 316–336.

Cole, D. A., & Maxwell, S. E. (2003). Testing mediational models with longitudinal data: Questions and tips in the use of structural equation modeling. *Journal of Abnormal Psychology*, *112*, 558–577.

Cole, P. M., & Kaslow, N. J. (1988). Interactional and cognitive strategies for affect regulation: Developmental perspective on childhood depression. In L. B. Alloy (Ed.), *Cognitive processes in depression* (pp. 310–343). New York: Guilford.

Cole, P. M., Martin, S. E., & Dennis, T. A. (2004). Emotion regulation as a scientific construct: Methodological challenges and directions for child development research. *Child Development*, *75*, 317–333.

Davies, P. T., & Windle, M. (1997). Gender-specific pathways between maternal depressive symptoms, family discord, and adolescent adjustment. *Developmental Psychology*, *33*, 657–668.

Denham, S. A. (1997). "When I have a bad dream, mommy holds me": Preschoolers' conceptions of emotions, parental socialisation, and emotional competence. *International Journal of Behavioral Development*, *20*, 301–319.

Denham, S. A., & Kochanoff, A. T. (2002). Parental contributions to preschoolers' understanding of emotion. *Marriage and Family Review*, *34*, 311–343.

Denham, S. A., Mitchell-Copeland, J., Strandberg, K., Auerbach, S., & Blair, K. (1997). Parental contributions to preschoolers' emotional competence: Direct and indirect effects. *Motivation and Emotion, 21*(1), 65 – 86.

Diamond, L. M., & Aspinwall, L. G. (2003). Emotion regulation across the life span: An integrative perspective emphasizing self-regulation, positive affect, and dyadic processes. *Motivation and Emotion, 27*(2), 125 – 156.

Eisenberg, N., Cumberland, A., & Spinrad, T. L. (1998). Parental socialization of emotion. *Psychological Inquiry, 9*, 241 – 273.

Eisenberg, N., & Fabes, R. A. (1994). Mothers' reactions to children's negative emotions—Relations to children's temperament and anger behavior. *Merrill-Palmer Quarterly-Journal of Developmental Psychology, 40*, 138 – 156.

Eisenberg, N., Fabes, R. A., & Murphy, B. C. (1996). Parents' reactions to children's negative emotions: Relations to children's social competence and comforting behavior. *Child Development, 67*, 2227 – 2247.

Eisenberg, N., Fabes, R. A., Shepard, S. A., Guthrie, I. K., Murphy, B. C., & Reiser, M. (1999). Parental reactions to children's negative emotions: Longitudinal relations to quality of children's social functioning. *Child Development, 70*, 513 – 534.

Ellis, L. K., & Rothbart, M. K. (2001). *Revision of the Early Adolescent Temperament Questionnaire*. Paper presented at the Biennial Meeting of the Society for Research in Child Development, Minneapolis, MN.

Fabes, R. A., Eisenberg, N., & Bernzweig, J. (1990). *The coping with Children's Negative Emotion Scale: Description and scoring*. Unpublished manuscript, Department of Resources and Human Development, Arizona State University, Tempe.

Field, T. (1994). The effects of mother's physical and emotional unavailability on emotion regulation. *Monographs of the Society for Research in Child Development, 59* (2 – 3, Serial no. 240), 208 – 227.

Fleiss, J. L. (1981). *Statistical methods for rates and proportions*. New York: Wiley.

Forbes, E. E., & Dahl, R. E. (2005). Neural systems of positive affect: Relevance to understanding child and adolescent depression? *Development and Psychopathology, 17*, 827 – 850.

Forgatch, M. S. (1989). Patterns and outcome in family problem-solving—The disrupting effect of negative emotion. *Journal of Marriage and the Family, 51*, 115 – 124.

Fox, N. A. (1994). The development of emotion regulation. *Monographs of the Society for Research in Child Development, 59* (2 – 3, Serial no. 240).

Fredrickson, B. L. (1998). What good are positive emotions? *Review of General Psychology, 2*, 300 – 319.

Garber, J., Braafladt, N., & Weiss, B. (1995). Affect regulation in depressed and nondepressed children and young adolescents. *Development & Psychopathology, 7*(1), 93 – 115.

Garber, J., Braafladt, N., & Zeman, J. (1991). The regulation of sad affect: An information-processing perspective. In J. Garber & K. A. Dodge (Eds.), *The development of emotion regulation and dysregulation. Cambridge studies in social and emotional development* (pp. 208 – 240). New York: Cambridge University Press.

Garber, J., & Dodge, K. A. (1991). *The development of emotion regulation and dysregulation*. New York: Cambridge University Press.

Gilboa, E. (1994). Personality and the structure of affective responses. In S. H. M. van Goozen, N. E. Van de Poll, & J. A. Sergeant (Eds.), *Emotions: Essays on emotion theory*. Hillsdale, NJ: Lawrence Erlbaum.

Gottman, J. M., Katz, L. F., & Hooven, C. (1996). Parental meta-emotion philosophy and the emotional life of families: Theoretical models and preliminary data. *Journal of Family Psychology, 10*, 243 – 268.

Gottman, J. M., Katz, L. F., & Hooven, C. (1997). *Meta-emotion: How families communicate emotionally*. Mahwah, NJ: Lawrence Erlbaum.

Gross, J. J. (1999). Emotion regulation: Past, present, future. *Cognition & Emotion, 13*, 551 – 573.

Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality & Social Psychology, 85*, 348 – 362.

Gross, J. J., & Levenson, R. W. (1997). Hiding feelings: The acute effects of inhibiting negative and positive emotion. *Journal of Abnormal Psychology, 106*, 95 – 103.

Gross, J. J., & Munoz, R. F. (1995). Emotion regulation and mental health. *Clinical Psychology: Science & Practice, 2*, 151 – 164.

Halberstadt, A. G., Crisp, V. W., & Eaton, K. L. (1999). Family expressiveness: A retrospective and new directions for research. In P. Philippot, R. S. Feldman, & E. J. Coats (Eds.), *The social context of nonverbal behavior* (pp. 109 – 155). Cambridge, UK: Cambridge University Press.

Hankin, B. L., Abramson, L. Y., Moffitt, T. E., Silva, P. A., McGee, R., & Angell, K. E. (1998). Development of depression from preadolescence to young adulthood: Emerging gender differences in a 10-year longitudinal study. *Journal of Abnormal Psychology, 107*, 128 – 140.

Hops, H., Davis, B., & Longoria, N. (1995). Methodological issues in direct observation—Illustrations with the Living in Familial Environments (LIFE) coding system. *Journal of Clinical Child Psychology, 24*, 193 – 203.

Hops, H., Sherman, L., & Biglan, A. (1990). Maternal depression, marital discord, and children's behavior: A developmental perspective. In G. R. Patterson (Ed.), *Depression and aggression in family interaction. Advances in family research* (pp. 185 – 208). Hillsdale, NJ: Lawrence Erlbaum.

Jacob, T., & Johnson, S. L. (1997). Parent-child interaction among depressed fathers and mothers: Impact on child functioning. *Journal of Family Psychology, 11*, 391 – 409.

Jacob, T., & Johnson, S. L. (2001). Sequential interactions in the parent-child communications of depressed fathers and depressed mothers. *Journal of Family Psychology, 15*, 38 – 52.

Keenan, K. (2000). Emotion dysregulation as a risk factor for child psychopathology. *Clinical Psychology: Science & Practice, 7*, 418–434.

Kessler, R. C., Avenevoli, S., & Merikangas, K. R. (2001). Mood disorders in children and adolescents: An epidemiologic perspective. *Biological Psychiatry, 49*, 1002–1014.

Klimes-Dougan, B., Brand, A. E., Zahn-Waxler, C., Usher, B., Hastings, P. D., Kendziora, K., et al. (2007). Parental emotion socialization in adolescence: Differences in sex, age and problem status. *Social Development, 16*, 326–342.

Kobak, R. R., Cole, H. E., Ferenz-Gillies, R., Fleming, W. S., & Gamble, W. (1993). Attachment and emotion regulation during mother-teen problem-solving: A control theory analysis. *Child Development, 64*, 231–245.

Ladouceur, C., Reid, L., & Jacques, A. (2002). Construction and validation of the Parents' Reaction to Children's Positive Emotions Scale. *Canadian Journal of Behavioural Science, 31*, 8–18.

Larsen, R. J. (2000). Toward a science of mood regulation. *Psychological Inquiry, 11*, 129–141.

Leinonen, J. A., Solantaus, T. S., & Punamaki, R. L. (2003). Parental mental health and children's adjustment: The quality of marital interaction and parenting as mediating factors. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 44*, 227–241.

Lewinsohn, P. M., Hops, H., Roberts, R. E., Seeley, J. R., & Andrews, J. A. (1993). Adolescent psychopathology 1: Prevalence and incidence of depression and other DSM–III–R disorders in high-school students. *Journal of Abnormal Psychology, 102*, 133–144.

Lindahl, K. M., & Markman, H. J. (1990). Communication and negative affect regulation in the family. In E. A. Blechman (Ed.), *Emotions and the family: For better or for worse* (pp. 99–115). Hillsdale, NJ: Lawrence Erlbaum.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods, 7*, 83–104.

MacKinnon, D. P., & Warsi, G. (1995). A simulation study of mediated effect measures. *Multivariate Behavioral Research, 30*, 41.

McFarlane, A. H., Bellissimo, A., & Norman, G. R. (1995). Family structure, family functioning and adolescent well-being: The transcendent influence of parental style. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 36*, 847–864.

Moos, R. H., & Moos, B. S. (1981). *Family Environment Scale and preliminary manual*. Palo Alto, CA: Consulting Psychologist Press.

Morris, A. S., Silk, J. S., Steinberg, L., Myers, S. S., & Robinson, L. R. (2007). The role of the family context in the development of emotion regulation. *Social Development, 16*, 361–388.

Preacher, K. J., & Leonardelli, G. J. (2001). Calculation for the Sobel test: An interactive calculation tool for mediation tests [Online software]. Retrieved February 2007 from http://quantrm2.psy.ohio-state.edu/kris/sobel/sobel.htm

Radloff, L. S. (1977). The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*, 385–401.

Radloff, L. S. (1991). The use of the Center for Epidemiologic Studies Depression Scale in adolescents and young adults. *Journal of Youth and Adolescence, 20*, 149–166.

Ramsden, S. R., & Hubbard, J. A. (2002). Family expressiveness and parental emotion coaching: Their role in children's emotion regulation and aggression. *Journal of Abnormal Child Psychology, 30*, 657–667.

Roberts, R. E., Lewinsohn, P. M., & Seeley, J. R. (1991). Screening for adolescent depression: A comparison of depression scales. *Journal of the American Academy of Child & Adolescent Psychiatry, 30*, 58–66.

Sheeber, L., Allen, N., Davis, B., & Katz, L. F. (2004). *National Institute of Health grant application*. Unpublished manuscript.

Sheeber, L., Allen, N., Davis, B., & Sorensen, E. (2000). Regulation of negative affect during mother-child problem-solving interactions: Adolescent depressive status and family processes. *Journal of Abnormal Child Psychology, 28*, 467–479.

Sheeber, L., Davis, B., Leve, C., Hops, H., & Tildesley, E. (2007). Adolescents' relationships with their mothers and fathers: Associations with depressive disorder and subdiagnostic symptomatology. *Journal of Abnormal Psychology, 116*, 144–154.

Sheeber, L., Hops, H., Alpert, A., Davis, B., & Andrews, J. (1997). Family support and conflict: Prospective relations to adolescent depression. *Journal of Abnormal Child Psychology, 25*, 333–344.

Sheeber, L., Hops, H., Andrews, J., Alpert, T., & Davis, B. (1998). Interactional processes in families with depressed and non-depressed adolescents: Reinforcement of depressive behavior. *Behaviour Research and Therapy, 36*, 417–427.

Silk, J. S., Steinberg, L., & Morris, A. S. (2003). Adolescents' emotion regulation in daily life: Links to depressive symptoms and problem behavior. *Child Development, 74*, 1869–1880.

Southam-Gerow, M. A., & Kendall, P. C. (2002). Emotion regulation and understanding: Implications for child psychopathology and therapy. *Clinical Psychology Review, 22*, 189–222.

Thayer, J. F., Rossy, L. A., Ruiz-Padial, E., & Johnsen, B. H. (2003). Gender differences in the relationship between emotional regulation and depressive symptoms. *Cognitive Therapy and Research, 27*, 349–364.

Tomarken, A. J., & Keener, A. D. (1998). Frontal brain asymmetry and depression: A self-regulatory perspective. *Cognition & Emotion, 12*, 387–420.

Vinson, T. (2004). *Community adversity and resilience: The distribution of social disadvantage in Victoria and NSW and the mediating role of social cohesion*. Richmond, Australia: Jesuit Social Services.

Yap, M. B. H., Allen, N. B., O'Shea, M., Di Parsia, P., Simmons, J. G., & Sheeber, L. (2007). *Early adolescents'* *temperament, emotion regulation during mother-child interactions and depressive symptomatology.* Manuscript submitted for publication.

Yap, M. B. H., Allen, N. B., & Sheeber, L. (2007). Using an emotion regulation framework to understand the role of temperament and family processes in risk for adolescent depressive disorders. *Clinical Child & Family Psychology Review, 10,* 180–196.

Youniss, J., & Ketterlinus, R. D. (1987). Communication and connectedness in mother- and father-adolescent relationships. *Journal of Youth & Adolescence, 16,* 265–280.

Appendix A: Parents' Reaction to Children's Positive Emotions Scale

Table A1

Intercorrelations Among Subscales of the Parents' Reaction to Children's Positive Emotions Scale

	Discomfort	Encouragement	Reprimand
Socialization	.38	−.25	.50
Discomfort		−.37	.65
Encouragement			−.46

Notes. N = 233, except the right-most column, where *N* = 232. All correlations are significant, *ps* < .001.

Examples of vignettes used (subscale are indicated in parentheses) are the following:

1. If we are sitting on the patio at our friends' house and, while playing with other children, my child is running, jumping and screaming, I would:
    (a) tell my child to calm down and suggest another game that he/she could play (Socialization)
    (b) tell my child, in a firm voice, to stop jumping and running around immediately (Reprimand)
    (c) let my child play and have fun (Encouragement)
    (d) be slightly embarrassed by my child's behavior (Discomfort).
2. If we are going on a family car-trip and, during the two-hour drive, my child and his/her friend are singing and laughing loudly, I would:
    (a) feel annoyed by my child's behavior (Discomfort)
    (b) allow my child to have fun with his/her friend (Encouragement)
    (c) firmly tell my child to stop it right now (Reprimand)
    (d) suggest a quiet game for them to play (Socialization).

Appendix B: Components of Living in Family Environments Constructs

Content codes	Meaning	Affect codes	Meaning
10	Validation	0	Contempt
14	Affection	1	Anger
15	Solicitous	2	Anxious
16	Humor	3	Dysphoric
18	Approve	4	Pleasant
21	Complaint	5	Neutral
23	Cruelty	6	Happy
24	Negative substance	7	Caring
25	Provoke	8	Whine
26	Annoy–disrupt	9	Belligerence
28	Disagree		
30	Command		
31	Command unaccountable		
35	Comply		
37	Noncomply		
46	Self-statement		
47	Self-positive		
48	Self-complaint		
61	Problem statement		
64	Propose solution		
65	Teach		
66	Guidance tactic		
76	Attend		
78	Talk		
81	Inaudible		
98	Dummy		

Notes. Aversive construct = Affect codes 0, 1, or 9+ any content code or Affect code 5+ content codes 23, 25, or 26. Dysphoric construct = Affect codes 2, 3, or 8+ any content code or Affect code 5+ content codes 21 or 48. Positive Interpersonal construct = Affect codes 4, 6, or 7+ any content code or Affect code 5+ content codes 10, 14, 16, 18, or 47.

Child Development, September/October 2008, Volume 79, Number 5, Pages 1432–1443

Infants' Use of Lexical-Category-to-Meaning Links in Object Individuation

D. Geoffrey Hall, Kathleen Corrigall, Mijke Rhemtulla,
Eleanor Donegan, and Fei Xu
University of British Columbia

Infants watched an experimenter retrieve a stuffed animal from an opaque box and then return it. This happened twice, consistent with either 1 animal appearing on 2 occasions or 2 identical-looking animals each appearing once. The experimenter labeled each object appearance with a different novel label. After infants retrieved 1 object from the box, their subsequent search behavior was recorded. Twenty-month-olds, but not 16-month-olds, searched significantly longer for a second object inside the box when the labels were both proper names than when they were 1 count noun followed by 1 proper name. The effect was not significant when proper names were replaced by adjectives. Twenty-month-olds' understanding of meaning distinctions among several word categories guided their object individuation.

By around 2 years, infants have learned that words from different lexical categories in their native language convey distinct types of meaning. Specifically, they have acquired the knowledge that novel words (e.g., "FEP") modeled linguistically as count nouns (e.g., "This is a FEP") label object categories, but those presented as members of other lexical categories carry different semantic implications. For example, infants of this age have discovered that words modeled as adjectives (e.g., "This is very FEP") designate object properties (e.g., Booth & Waxman, 2003; Waxman, 1999; Waxman & Booth, 2001; Waxman & Markow, 1998), words modeled as verbs (e.g., "This is FEPing") mark action categories (e.g., Echols & Marti, 2004), words modeled as prepositions (e.g., "This is FEP my box") label spatial locations (e.g., Fisher, Klingler, & Song, 2006), and words modeled as proper names (e.g., "His name is FEP") refer to individual objects (e.g., Bélanger & Hall, 2006; Hall, Lee, & Bélanger, 2001).

Knowledge of lexical-category-to-meaning links gives infants a valuable tool for constraining the interpretation of novel words. As a result, it significantly influences categorization behavior in a variety of word extension tasks. For example, in a study of infants' understanding of the distinction between count nouns and proper names, Bélanger and Hall (2006) taught infants a novel label for a stuffed animal. One group heard the word modeled as a count noun (e.g., "He is called a DAX"), and another group heard the same word modeled as a proper name (e.g., "He is called DAX"). The animal was then moved to a new location in front of the infant, and a second same-looking animal was placed in the first animal's original location. Infants were asked to look at one of the two objects as a referent of the novel word, modeled either as a count noun (e.g., "Find a DAX") or as a proper name (e.g., "Find DAX"). At 16 months, infants' looking patterns were unaffected by the word's lexical category. At 20 months, however, infants looked first to the labeled animal significantly more often after hearing a proper name than after hearing a count noun (Bélanger & Hall, 2006; for further evidence, see Bélanger & Hall, 2008; Hall et al., 2001; Liittschwager & Markman, 1993; Sorrentino, 2001; Sorrentino & Halberda, 2001). By 20 months, infants thus appeared to have learned lexical-category-to-meaning links for proper names and count nouns in their language, and this knowledge guided their interpretation of novel words.

Another fundamental task facing infants during conceptual development is object individuation. How do infants make appropriate decisions regarding the number of distinct objects present in a scene? Recent research suggests that labeling can influence these judgments. For example, Xu, Cote, and Baker (2005) presented 12-month-old infants with an opaque box

---

This work was funded by an operating grant to D. Geoffrey Hall from the Natural Sciences and Engineering Research Council of Canada. Portions of the research were presented at the Boston University Conference on Language Development, November 2006. We thank Cristy McNiven for her help with data collection.

Correspondence concerning this article should be addressed to D. Geoffrey Hall, Department of Psychology, University of British Columbia, 2136 West Mall, Vancouver, BC, Canada, V6T 1Z4. Electronic mail may be sent to geoff@psych.ubc.ca.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7905-0014

into which they could reach but not see. In the two-word condition, the experimenter looked into the box and said, "Look, a FEP! Look, a WUG!" In the one-word condition, she said, "Look, a ZAV! Look, a ZAV!" After they retrieved one object from inside the box, infants were allowed to reach again. The duration of this second search was the dependent measure of interest, and because the box was empty, search time was interpreted as reflecting the strength of infants' expectation that the box contained two objects. Xu et al. found that 12-month-olds searched significantly longer in the two-word than in the one-word condition, suggesting that after hearing two different count nouns, infants inferred that two objects were hidden in the box. The finding also offers evidence that infants at this age expect that two different count nouns should not apply to one object, perhaps an early manifestation of the assumption of mutual exclusivity (Markman, 1989).

The previous work by Xu et al. (2005) focused on the effect of hearing count noun labels on object individuation among infants who have not yet acquired knowledge of multiple lexical-category-to-meaning links in their language. In the current research, we investigated labeling effects among older infants who have learned at least some of these links. We used a modified version of Xu et al.'s task (cf. Feigenson & Carey, 2003; Van de Walle, Carey, & Prevor, 2000; Xu & Baker, 2005) to explore whether these older infants can use their understanding of the distinct semantic implications of count nouns and proper names to individuate objects.

In our task, we first administered training trials on which infants saw that an opaque box could contain either one or two identical-looking stuffed animals. Infants then took part in two test trials. On each test trial, the experimenter showed infants a new opaque box, retrieved a stuffed animal from inside it and then placed the animal back inside. The experimenter then repeated these actions, retrieving the same animal from inside the box and returning it. Because the two animal appearances looked identical, the event was ambiguous: It could have involved the same animal being shown twice or two different but identical-looking animals, each being shown once. As a result, it was possible to infer either that there was one animal or that there were two distinct but same-looking animals inside the box. Each time she retrieved the animal, the experimenter labeled it with a different novel word. On one test trial, infants heard a different proper name for each appearance of the animal (e.g., "His name is FEP"; "His name is BLICK"), and on the other trial, they heard a count noun followed by a proper name for the two appearances of a stuffed animal of a different kind (e.g., "He is a MUV"; "His name is PRAZ"). On each test trial, the experimenter then invited infants to search for what was inside the box. After retrieving the sole stuffed animal that was inside, infants were allowed to search again. As in Xu et al. (2005), our dependent measure was the duration of this second search, interpreted as a reflection of how strongly infants expected that the box contained two objects.

A test trial in our task had several distinctive features. To begin, it differed from a test trial in Xu et al.'s (2005) task in three primary ways. First, the experimenter showed an object to accompany each novel word, unlike Xu et al.'s task, in which the experimenter simply looked into the box when providing the labels. Second, the experimenter provided a different novel label for each object emergence, unlike Xu et al.'s task, in which the experimenter provided different novel words on one trial but repeated a single novel word on the other. Third, the experimenter varied the lexical category of these labels, unlike Xu et al.'s task, in which the experimenter presented all words as count nouns. Note also that infants in our task could not use spatiotemporal information to resolve the ambiguity surrounding the number of objects present in the box because they saw only one object appearance at a time. Similarly, they could not use property or kind information because the two object appearances had the same properties and were of the same kind. Infants could, however, potentially resolve the ambiguity by using the lexical category information accompanying each object appearance. The linguistic information differed between test trials, and for this reason, we interpreted differences in the duration of the second search on the two trials as reflecting infants' use of these lexical category cues to disambiguate the event.

If infants understand that novel count nouns designate object categories, whereas proper names mark individual objects, we expected that they would individuate the event differently on the two test trials. On the two proper names trial, we predicted that they would infer the presence of two stuffed animals inside the box (i.e., an animal named FEP and another same-looking animal named BLICK) because they would judge two proper names for one object to be a violation of the contrast principle (i.e., different words have different meanings; see Clark, 1993). If both words referred to the same object, they would both provide a name for the same individual object, in violation of the principle. As a result, we expected that infants would reject this possibility (see Hall & Graham, 1999). On the count noun – proper name trial, however, we anticipated that infants would be more likely to infer that the box contained only one object (i.e., an animal that was a MUV, whose name was PRAZ) because they would judge that one count noun and one proper name for one object would not violate contrast (i.e., the two words would have different meanings, one an object category term and the other a name for the individual object). As a result, we expected that they would allow the possibility that this pair of words applied to the same object. After they retrieved the single object inside the box, we thus expected that infants would search more persistently (i.e., longer) on the two proper names trial than on the count noun–proper name trial.

In Study 1, we tested 20-month-olds using sentences that contained multiple cues to distinguish proper names from count nouns. We expected that these infants would know the semantic distinction between count nouns and proper names (cf. Bélanger & Hall, 2006) and, as a result, would individuate the event differently on the two trials. Study 2 was a modified replication of Study 1, in which the sentences containing the novel words held fewer cues to distinguish between count nouns and proper names. Study 3 was a replication of Study 1 with 16-month-olds. We predicted that they would not individuate the event differently on the two trials (cf. Bélanger & Hall, 2006). Studies 4 and 5 examined the basis for 20-month-olds' success on our task. Together, the studies provide the first assessment in the literature of whether infants' nascent understanding of distinct lexical-category-to-meaning links guides them in the fundamental task of object individuation.

Study 1

Method

Participants. Participants were 16 infants (8 girls and 8 boys), with an average age of 20.3 months (range = 18.5–22.5 months). We also tested, but excluded from the final sample, 5 infants for failing to complete the task, and 1 for having a search time longer than 2.5 *SD* from the mean on at least one trial. In this study and subsequent studies, all infants were acquiring English as their native language, and most were from middle- or upper-middle-class Caucasian families. Infants were recruited from a database maintained by a developmental research group in a university psychology department. Infants received a T-shirt and a certificate for participating; parents were reimbursed for parking or transit expenses.

Materials. On training trials, we used one stuffed dog and two identical stuffed monkeys. On test trials, we used one stuffed bear and one stuffed rabbit. All were between 12 and 18 cm in their longest dimension.

We also used one opaque box on training trials (44 × 30 × 21 cm) and two others on test trials (29 × 25 × 21 cm and 29 × 25 × 23 cm). Each had an opening on one side (roughly 17 × 10 cm), covered in fabric strips. As a result, infants could reach but not see inside.

The study was conducted at a table in a playroom of a developmental laboratory in a university psychology department. The experimenter sat at a 90° angle to the infant on the infant's left side. The infant sat either in a booster chair or on her parent's lap. A digital video camera recorded the infant's search behavior and was placed so that it faced the experimenter and captured the infant's right side profile.

Design and procedure. All infants took part in two training trials followed by two test trials.

Training trials. The purpose of the training trials was to ensure that infants understood that a box could contain either one or two stuffed animals, and to give them experience in reaching inside. On the first trial, the experimenter placed the empty training box on the table in front of the infant, put the stuffed dog inside, and asked the infant, "Can you get what's inside?" If the infant did so, the experimenter offered praise, took the toy, and thanked her. Keeping the stuffed dog in the infant's view, the experimenter pushed the box back toward her and asked, "Do you think there's anything else inside the box?" The infant then received a second opportunity to search. Regardless of the infant's response, the experimenter took the box after 10 s, shook it, and said, "It's empty!"

The second trial was similar to the first, except that the experimenter placed the two identical-looking stuffed monkeys inside the empty training box. If the infant retrieved one toy on each of the two reach opportunities, the experimenter offered praise, took the toy, and thanked the infant each time. If the infant retrieved both toys on the first reach opportunity, the experimenter offered praise, took the toys, and thanked the infant.

On either training trial, if the infant failed to retrieve a toy that was still inside the box, the experimenter provided assistance by pulling a small part of the object through the opening. Any infant who still failed to retrieve a toy on either training trial was excluded from the study.

Proper name–proper name (PN-PN) test trial. The experimenter placed one of the two test boxes on the table. The box already contained either the stuffed bear or the stuffed rabbit. The experimenter reached into the box, pulled out the toy, and repeated a novel proper name five times. She said, "Look! His name is

FEP! Do you see FEP? I think FEP wants to come see you! Do you want to play with FEP?" The infant was allowed to play with the toy for approximately 10 s. The experimenter then took back the toy and thanked the infant. The experimenter said, "I'm going to put FEP back inside the box!" and placed the toy back inside.

The experimenter then reached back into the box and repeated the procedure, using a second proper name, "BLICK."

Count noun–proper name (CN-PN) test trial. The experimenter placed the other test box on the table. It already contained the stuffed animal not used on the PN-PN trial. The procedure was the same as on the PN-PN trial except that the experimenter presented the first novel word as a count noun. The experimenter said, "Look! He is a MUV! Do you see the MUV? I think the MUV wants to come see you! Do you want to play with the MUV?" Following the 10-s play period, the experimenter said, "I'm going to put the MUV back inside the box!"

The experimenter then reached back into the box and repeated the procedure using a second word, "PRAZ," introduced as a proper name, exactly as on the PN-PN trial.

Note that the only linguistic difference between test trials was in the sentence frames used to introduce the first novel word. On the PN-PN trial, the first sentence frame containing the first novel word (a proper name) was "His name is ___," and the novel word in the four subsequent sentence frames had no determiner. On the CN-PN trial, the first sentence frame containing the first novel word (a count noun) was "He is a ___," and the determiner, "the," preceded the novel word in the four subsequent sentence frames.

On both test trials, the experimenter then pushed the box toward the infant, with the covered opening in front of the infant and asked, "Can you get what's inside the box?" The infant was required to retrieve the single animal inside. If the infant did not do so, the experimenter provided assistance by pulling a small part of the object through the opening. Any infant who still failed to retrieve the toy was excluded from the study. After retrieving the toy, the infant was allowed to play with it for approximately 10 s.

The experimenter then took the toy back, placed it in the infant's view, pushed the box back toward the infant, and gave the infant a second opportunity to search inside. The duration of this second search was our dependent variable. It is important to note that there was never a second object in the box on test trials. It is also important to note that infants never received any encouragement or assistance on this second search opportunity. If infants did not initiate a search within 10 s, the trial ended.

We counterbalanced between participants both the order of the two test trials (PN-PN and CN-PN) and the order in which we presented the two kinds of toys (bear and rabbit), ensuring that each kind of toy was used half the time on each type of test trial. We also counterbalanced between participants the order in which we presented the two words associated with each animal (i.e., FEP and BLICK for the bear; "MUV" and "PRAZ" for the rabbit).

Data analysis. A primary coder, blind to trial order and trial type, coded the videotapes frame by frame for the duration of infants' second search. As in previous research using this paradigm (Van de Walle et al., 2000; Xu et al., 2005), we considered an infant to be searching inside the box when the knuckles that join the fingers to the hand were inside the box's opening. A second coder, also blind to trial order and trial type, subsequently recoded the data. Intercoder agreement was calculated to be $r = .99$. Trials on which there was a disagreement of 0.5 s or more between coders were reviewed by both coders together, discussed, and subsequently retimed by the primary coder.

Results and Discussion

As predicted, 20-month-olds searched longer on the PN-PN trial than on the CN-PN trial. We conducted a 2 × 2 mixed analysis of variance (ANOVA) on the search duration data, with trial order (PN-PN first, CN-PN first) as the between-subjects factor and trial type (PN-PN, CN-PN) as the within-subjects factor. The only significant effect was that for trial type, $F(1, 14) = 10.73$, $p = .006$, partial $\eta^2 = .434$ (Figure 1). In addition, 12 of the 16 infants' second

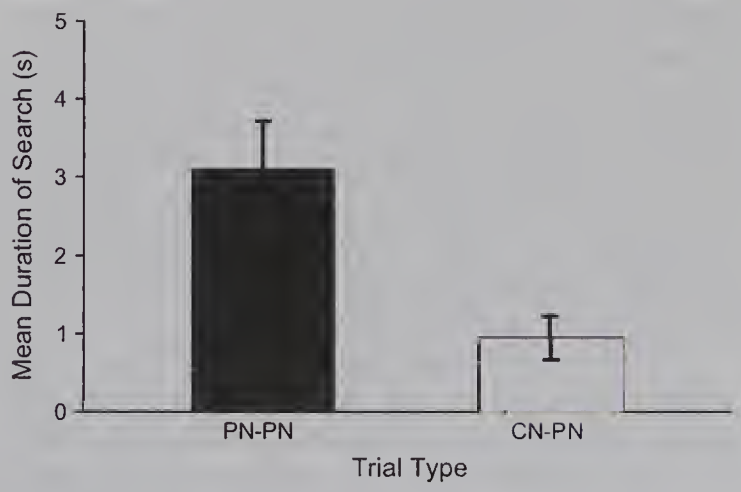


Figure 1. Results of Study 1.
Note. Bars represent 1 *SE*. PN = proper name; CN = count noun.

searches were longer on the PN-PN trial than on the CN-PN trial, $p = .038$ (one tailed) by a sign test.

Twenty-month-olds searched more persistently for a second stuffed animal inside a box after observing two appearances of the same animal labeled with different proper names than after observing them labeled with a count noun followed by a proper name. The result suggests that 20-month-olds distinguished between proper names and count nouns, and it indicates that they were more likely to expect two objects to be in the box after hearing two proper names than after hearing the combination of a count noun followed by a proper name. The finding also suggests that infants' nascent knowledge of the semantics of different lexical categories guided their object individuation.

The results are consistent with the claim that 20-month-olds understand that proper names label individual objects, whereas count nouns label object categories. The findings also suggest that infants have knowledge of conditions under which two novel words should pick out two objects, such as when the words would violate the contrast principle if they referred to the same object (Clark, 1993). If infants understand these lexical-category-to-meaning links and honor contrast, then we would expect them to reject two proper names for the same object because the two words would have the same meaning (i.e., they would both refer to the same individual object). We would not, however, expect infants to reject the combination of one count noun followed by one proper name for one object because this word combination would not violate contrast (i.e., the words would have different meanings; one would refer to an object category, whereas the other would refer to an individual object; see Clark, 1993; see also Diesendruck, 2005; Hall & Graham, 1999).

We note that our interest was in determining whether there were *any* circumstances under which infants would distinguish a trial involving two proper names (i.e., a PN-PN trial) from a trial involving one count noun and one proper name. We obtained our positive evidence using a trial on which the count noun *preceded* the proper name (i.e., a CN-PN trial). We make no claim about 20-month-olds' ability to distinguish a PN-PN trial from a PN-CN trial, although we have no a priori reason to expect that they would be unable to do so.

In Study 1, infants received multiple linguistic cues to help them distinguish proper names from count nouns. The sentences containing the first mention of the terms differed by three words (i.e., "His name is ___" vs. "He is a ___"). The sentences containing the subsequent four mentions of the word differed in terms of the absence or presence of the determiner, "the." However, recent research (Bélanger & Hall, 2006, 2008) has found evidence that 20-month-olds can distinguish semantically between proper names and count nouns, given only the presence or absence of the determiner as a cue. In Study 2, we sought to replicate the results of Study 1, giving infants only this minimal cue.

Study 2

Method

Participants. Participants were 16 infants (7 girls and 9 boys), with an average age of 20.7 months (range = 19.0–22.3 months). None had taken part in Study 1. We also tested, but excluded from the final sample, 1 infant for failing to complete the task and 2 for having a search time longer than 2.5 *SD* from the mean on at least one trial.

Materials. These were the same as in Study 1.

Design and procedure. These were the same as in Study 1 except that the sentence frames containing the first mention of the proper name and the count noun now differed only in terms of the presence or absence of a determiner ("a" or "the") before the novel word. Proper names were first introduced in the sentence, "This is ___", and count nouns were introduced in the frame, "This is a ___." Note that this change in wording makes the task of identifying the proper names more challenging because the frame, "This is ___," leaves open the possibility that the novel word is a proper name (e.g., "This is John"), an adjective (e.g., "This is pink"), or a mass noun (e.g., "This is fur"). As a result, infants in Study 2 needed to rely on the subsequent four sentences to establish unambiguously that the novel word was a proper name (e.g., "I think X wants to come see you" is felicitous only if "X" is a proper name).

Data analysis. Coding was carried out as in Study 1. Intercoder agreement was calculated to be $r = .95$.

Results and Discussion

As in Study 1, 20-month-olds searched longer on the PN-PN trial than on the CN-PN trial. We conducted the same ANOVA as in Study 1. The only significant effect was that for trial type, $F(1, 14) = 5.28$, $p = .038$, partial $\eta^2 = .274$ (Figure 2). Unlike Study 1, however, only 10 of the 16 infants' second reaches were longer on the PN-PN trial than on the CN-PN trial, nonsignificant by a sign test.

As in Study 1, 20-month-olds appeared to distinguish between proper names and count nouns, and the findings again suggest that infants were more

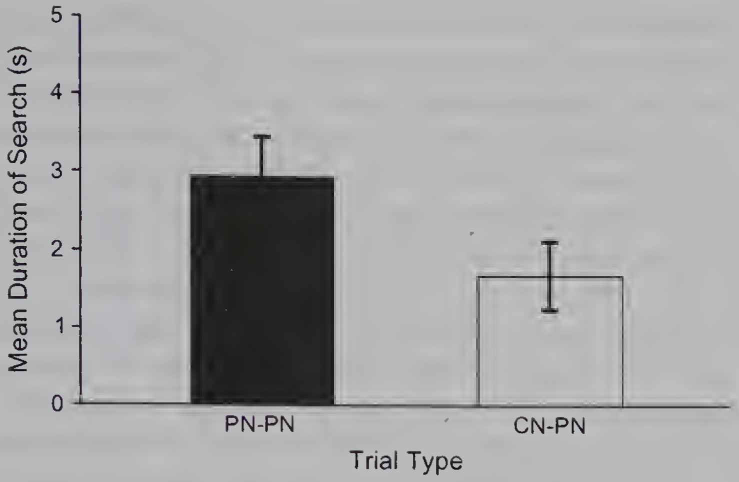


Figure 2. Results of Study 2.
Note. Bars represent 1 *SE*. PN = proper name; CN = count noun.

likely to expect two objects in the box after hearing two proper names than after hearing the combination of a count noun followed by a proper name. However, the trial type effect size was smaller in this study (partial $\eta^2 = .274$) than it was in Study 1 (partial $\eta^2 = .434$), as was the number of infants searching longer on the PN-PN trial (10 in Study 2 vs. 12 in Study 1). This difference between studies suggests that infants' ability to distinguish between the two lexical categories was weaker when they received only the presence or absence of a determiner as a cue to the distinction.

To analyze the effect of the number of linguistic cues on the effect of trial type, we combined the data from Studies 1 and 2 and conducted a 2 × 2 × 2 mixed ANOVA, with number of cues (multiple in Study 1 and minimal in Study 2) and trial order (PN-PN first, CN-PN first) as between-subjects factors, and trial type (PN-PN, CN-PN) as a within-subjects factor. There were no interactions involving number of cues, suggesting that despite any apparent differences between the results of the two studies, 20-month-olds distinguished comparably between proper names and count nouns in both.

Study 3 was an attempt to replicate the results of Study 1 with 16-month-olds. In previous research, children of this age have failed to distinguish between proper names and count nouns in a categorization task involving word extension (Bélanger & Hall, 2006; see also Bélanger & Hall, 2008). As a result, we predicted that 16-month-olds would search equivalently on the two test trials of our individuation task.

Study 3

Method

Participants. Participants were 16 infants (8 girls and 8 boys), with an average age of 16.5 months (range = 15.4 – 17.8 months). We also tested, but excluded from the final sample, 7 infants for failing to complete the task and 1 for having a search time longer than 2.5 *SD* from the mean on at least one trial.

Materials. These were the same as in Study 1.

Design and procedure. These were the same as in Study 1.

Data analysis. Coding was carried out as in Study 1. Intercoder agreement was calculated to be $r = 1.0$.

Results and Discussion

Unlike Study 1, 16-month-olds did not search significantly longer on the PN-PN trial than on the CN-PN trial. We conducted the same ANOVA as in Study 1. There were no significant effects, including that for trial type, $F(1, 14) = 0.23$, $p = .64$, partial $\eta^2 = .016$ (Figure 3). Only 6 of the 16 infants' searches were longer on the PN-PN trial than on the CN-PN trial, nonsignificant by a sign test.

Why did 16-month-olds not search significantly longer on the PN-PN trial than on the CN-PN trial? One possibility is that 16-month-olds simply fail at this type of search task. This alternative can be ruled out by prior demonstrations that infants as young as 10 – 12 months succeed in similar search tasks when given spatiotemporal or property/kind information (e.g., Feigenson & Carey, 2003; Van de Walle et al., 2000; Xu & Baker, 2005). Another possibility is that 16-month-olds are unable to use verbal information in search tasks. This alternative can be ruled out by the findings of Xu et al. (2005), revealing that infants as young as 12 months can use verbal information successfully in a similar search task. Given these considerations, we interpret the failure in Study 3 as reflecting the fact that, unlike 20-month-olds, 16-month-olds did not distinguish semantically

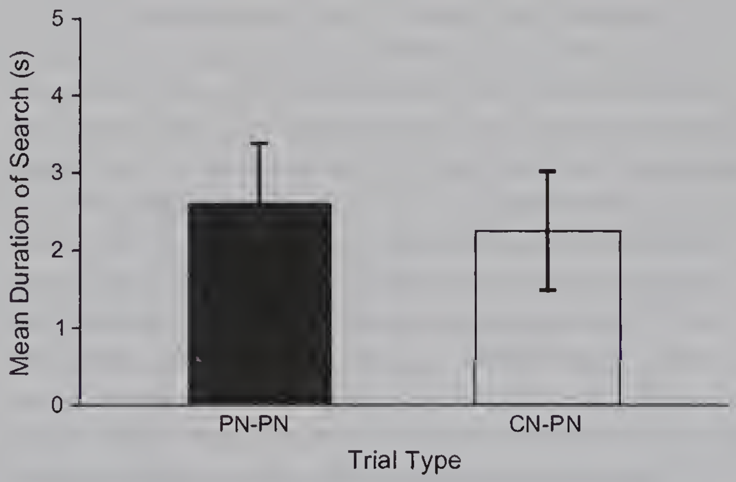


Figure 3. Results of Study 3.
Note. Bars represent 1 *SE*. PN = proper name; CN = count noun.

between proper names and count nouns and as a result were no more likely to expect two objects after hearing two proper names than after hearing the combination of a count noun followed by a proper name.

The results of Study 3 (with 16-month-olds) differed clearly from those of Study 1 (with 20-month-olds). In Study 3, the effect of trial type was not significant, the effect size was small (partial $\eta^2 = .016$), and a nonsignificant number of infants (6/16) showed the effect. In Study 1, recall that the effect of trial type was significant, the effect size was larger (partial $\eta^2 = .434$), and a significant number of infants (12/16) showed the effect. To follow up these separate analyses, we performed a direct between-study comparison. We combined the search duration data from Studies 1 and 3 and conducted a 2 × 2 × 2 mixed ANOVA, with age (16-month-olds in Study 3; 20-month-olds in Study 1) and trial order (PN-PN first, CN-PN first) as between-subjects factors and trial type (PN-PN, CN-PN) as a within-subjects factor. In this follow-up analysis, there was a main effect of trial type, $F(1, 28) = 6.23$, $p = .019$, partial $\eta^2 = .182$, but the expected interaction between trial type and age failed to reach significance, $F(1, 28) = 3.16$, $p = .086$, partial $\eta^2 = .101$. The fact that this interaction approached but did not attain significance is consistent with the possibility that sensitivity to the semantic distinction between proper names and count nouns is starting to emerge at 16 months.

These findings converge with previous results from Bélanger and Hall (2006), who found that 16-month-olds, unlike 20-month-olds, failed to distinguish between proper names and count nouns in a categorization task involving word extension. Note that Bélanger and Hall's task implicated very different abilities from those implicated by the current task: Their task assessed whether infants can extend novel proper names and count nouns appropriately; the current task, in contrast, assessed whether infants can use the proper name–count noun distinction to determine the number of objects present in a situation in which numerical identity is ambiguous. The agreement between the results from the two tasks thus lends further support to the proposal that infants acquire knowledge of the distinct semantics of these lexical categories between 16 and 20 months.

Study 4 was an attempt to clarify the basis for the trial type effect uncovered in Study 1. In Study 1, the second novel word was presented in the same sentence frames on both trials (i.e., as a proper name). However, the first novel word was presented in different sentence frames on the two trials, depending on whether it was a proper name or a count noun. We have suggested that infants relied on the difference between the sentence frames carrying the first word on the two trials to guide them in our task. However, we were also aware of a prosodic difference between trials associated with the sentence frames containing the *second* novel word. On the PN-PN trials (where the previous word had also been a proper name), the experimenter tended to place stress on the word "his" in the initial sentence frame, "His name is ___"; however, on the CN-PN trials (where the previous word had been a count noun), she tended to place\ emphasis on the word "name" in the same sentence frame. We wondered whether infants relied on this prosodic difference to guide their interpretations.

We began to address this issue by seeking evidence that adults could detect a prosodic difference between the initial sentence frames containing the second word when presented on the two trials. We obtained the audio clips of the presentation of this sentence frame from both trials from all 16 children in Study 1. This yielded 32 clips, each involving the presentation of a novel proper name. For 16 of the recordings, the word had been preceded by another proper name (i.e., on the PN-PN trial); for the other 16, the preceding word had been a count noun (i.e., on the CN-PN trial). Four adults listened to all these clips in a random order and decided, for each one, whether they heard stress on the word "his" or on the word "name."

We scored each adult's set of responses out of 32 to reflect their accuracy in grouping the recordings into those from the PN-PN trial and those from the CN-PN trial. A score of 50% (16/32) was taken to be random performance, and scores higher than 50% were taken to reflect successful categorization (i.e., judging stress on "his" on a PN-PN trial). All four adults were perfect (32/32) in their grouping. They thus detected a prosodic difference between the initial sentence frames containing the second word on the two trials in Study 1.

Our next step was to determine whether the prosodic difference in Study 1 contributed to the trial type effect discovered in that study. To do so, we conducted Study 4 as a replication of Study 1 with one change: We presented the second word with a similar prosody on the two trials.

Study 4

Method

Participants. Participants were 16 infants (8 girls and 8 boys), with an average age of 20.7 months (range = 18.0–22.4 months). None had taken part in any previous study. We also tested, but excluded from the final sample, 7 infants for failing to complete the task and 2 for having a search time longer than 2.5 *SDs* from the mean on at least one trial.

Materials. These were the same as in Study 1.

Design and procedure. These were the same as in Study 1 except that the experimenter was trained to present the second word on the two trials using similar prosody.

Data analysis. Coding was carried out as in Study 1. Intercoder agreement was calculated to be $r = 1.0$.

Before analyzing the data, we conducted a manipulation check. We enlisted the same four adults who coded the prosody of the initial sentence frames containing the second word for the Study 1 data to partition clips of these sentence frames for the Study 4 data, given the same directions. Recall that for the Study 1 data, all four adults had perfectly divided the clips into those from the PN-PN trial and those from the CN-PN trial (i.e., the average score was 100%). In contrast, adults in Study 4 were essentially random in the accuracy of their grouping. The average accuracy score was 51% (15/32, 16/32, 17/32, and 17/32). This result suggests that adults did not detect a systematic prosodic difference in the way the initial sentence frames containing the second word were presented on the two trials. This finding does not imply that we eliminated all prosodic differences or that infants similarly found no systematic prosodic difference between the sentences containing the second words on the two trials. However, the finding does suggest that the prosodic difference between trials identified in Study 1 was significantly reduced in Study 4.

Results and Discussion

As in Study 1, 20-month-olds searched longer on the PN-PN trial than on the CN-PN trial. We conducted the same ANOVA as in Study 1. The only significant effect was that for trial type, $F(1, 14) = 7.24$, $p = .018$, partial $\eta^2 = .341$ (Figure 4). As in Study 1, 12 of the 16 infants' second reaches were longer on the PN-PN trial than on the CN-PN trial, $p = .038$ (one tailed) by a sign test.

Despite the fact that we controlled the prosodic information provided with the second word on the two trials, 20-month-olds still appeared to distinguish between proper names and count nouns, and they again seemed more likely to expect two objects in the box when given two proper names than when given the combination of a count noun followed by a proper name. This result suggests that the prosodic difference identified by adults in the data from Study 1 was not the source of the trial type effect observed in that study; instead, the finding suggests that the source of the effect was the difference between the sentence frames in which the proper names and count nouns appeared.

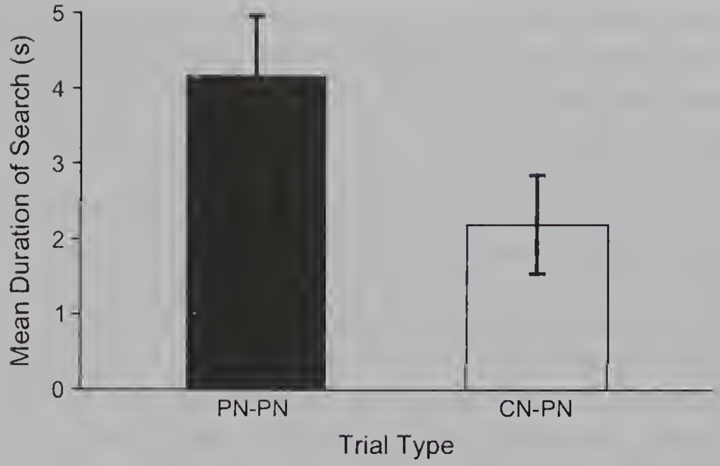


Figure 4. Results of Study 4.
Note. Bars represent 1 *SE*. PN = proper name; CN = count noun.

Study 5 was an attempt to clarify further the effect of trial type observed among 20-month-olds. As noted in Study 1, one interpretation is that the effect stems from infants' understanding of the distinct semantics of proper names and count nouns, coupled with adherence to the contrast principle. An alternative, however, is that the effect reflects infants' understanding of the pragmatic implications of a speaker's using two novel object labels from the same lexical category rather than two labels from different lexical categories. On this alternative, infants infer that a speaker who uses two novel labels from the same lexical category likely intends to use them contrastively to refer to two different objects, whereas they draw no such inference about a speaker who uses two words from different lexical categories. On this alternative, the trial type effect is not due to specific knowledge about the distinct semantics of proper names and count nouns.

In Study 5, we tested between these alternatives by replicating Study 4 but replacing proper name sentence frames with adjective sentence frames. Unlike proper names, which refer to individual objects, adjectives refer to object properties. Recent research suggests that infants have some understanding of the link between adjectives and object properties by as young as 14 months (e.g., Booth & Waxman, 2003; Waxman, 1999). If the findings we obtained with 20-month-olds in our previous studies reflect infants' inferences about a speaker's intended use of two words from the same lexical category rather than two words from different lexical categories, then we should have replicated the significant trial type effect observed in Study 4. In other words, we should have found that the trial type effect that emerged when we compared two proper names to the combination of one count noun followed by one proper name also emerged equivalently when we compared two adjectives to the combination of one count noun followed by one adjective.

If, however, the effect observed with 20-month-olds in our previous studies reflects infants' knowledge of the distinct semantics of proper names and count nouns, coupled with adherence to the contrast principle, then the trial type effect should have been *weaker* in Study 5. To understand this expectation, note that the contrast principle makes different predictions about the interpretation of adjectives in our task than it does about the interpretation of proper names. Allowing two proper names for one object would violate the contrast principle because the words would have the same meaning (i.e., they would both be labels for the same individual object). Allowing the combination of one count noun followed by one proper name would *not* violate contrast, however, because the words would have different meanings (i.e., one would label an object category and the other would label an individual object). As a result, infants in our task should reject two proper names for one object but be willing to accept the combination of one count noun followed by one proper name. This prediction was supported by the significant trial type effect observed with 20-month-olds in the previous studies.

If adjectives replace proper names in our task, however, the prediction changes. Allowing two novel adjectives for one object would *not* necessarily violate the contrast principle because the words could have different meanings (i.e., they could be labels for different object properties). Similarly, allowing one count noun followed by one adjective would not violate the contrast principle because the two words would have different meanings (i.e., one would label an object category and the other would label an object property). As a result, infants should be willing to accept both the possibility that two adjectives apply to the same object as well as the possibility that the combination of one count noun followed by one adjective applies to that object (see also Diesendruck, Hall, & Graham, 2006; Hall & Graham, 1999). The trial type effect that we observed with 20-month-olds in our previous studies should thus be weaker when we replace proper names with adjectives in our task.

Study 5

Method

Participants. Participants were 16 infants (8 girls and 8 boys), with an average age of 20.6 months (range = 18.1 – 22.5 months). None had taken part in any previous study. We also tested, but excluded from the final sample, 4 infants for failing to complete the task and 2 for having a search time longer than 2.5 *SDs* from the mean on at least one trial.

Materials. These were the same as in Study 4.

Design and procedure. These were similar to those used in Study 4, except that the novel words modeled as proper names in Study 4 were now modeled as adjectives. As a result, the two trials were named the adjective – adjective (ADJ-ADJ) trial and the count noun – adjective (CN-ADJ) trial. The five sentence frames we used to present the adjectives were similar to those used to present proper names in Study 4: "Look! He is really FEP! Do you see? He is very FEP! Look! He is very FEP! I think he wants to come see you! Do you want to play with him? He is very FEP! Look, he is very FEP! I'm going to put him back inside the box!"

Data analysis. Coding was carried out as in Study 1. Intercoder agreement was calculated to be $r = .99$.

As in Study 4, we conducted a manipulation check to determine whether adults detected any systematic prosodic difference between the sentence frames containing the first mention of the second word on the two trials (i.e., "He is really ___"). As in Study 4, the four adult raters were essentially random in the accuracy of their partitioning of the 32 recordings into those that were from the ADJ-ADJ trial (i.e., judged stress on "he") and those that were from the CN-ADJ trial (i.e., judged no stress on "he"). The average accuracy score was 50% (18/32, 18/32, 13/32, and 15/32). This finding suggests that adults did not detect any systematic prosodic difference between the initial sentence frames containing the second novel word on the two trials.

Results and Discussion

Twenty-month-olds did not search significantly longer on the ADJ-ADJ trial than on the CN-ADJ trial. We conducted the same ANOVA as in Study 1. There were no significant effects, including that for trial type, $F(1, 14) = 1.38$, $p = .26$, partial $\eta^2 = .089$ (Figure 5). Only 8 of the 16 infants' searches were longer on the ADJ-ADJ trial than on the CN-ADJ trial, nonsignificant by a sign test.

When we replaced proper names with adjectives in the task from Study 4, 20-month-olds did not search significantly more persistently after hearing two novel words from the same lexical category (two adjectives) than after hearing two words from different lexical categories (one count noun followed by one adjective). This discovery suggests that the significant

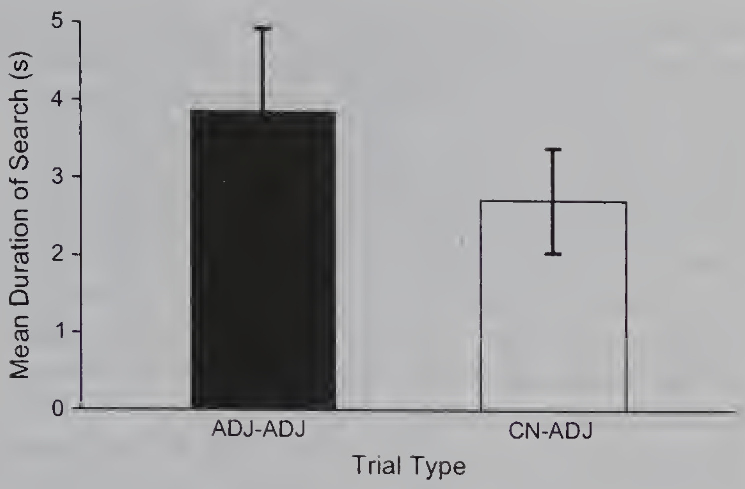


Figure 5. Results of Study 5.
Note. Bars represent 1 *SE*. ADJ = adjective; CN = count noun.

trial type effect observed with 20-month-olds in Study 4 was not driven solely by our use of two words from the same lexical category rather than two words from different lexical categories. Instead, the result offers support for the claim that the effect reflected infants' knowledge of the distinct semantics of count nouns and proper names, coupled with adherence to the principle of contrast.

The results of Study 5 (with adjectives) clearly differed from those of Study 4 (with proper names). In Study 5, the trial type effect was not significant, the effect size was modest (partial η^2 = .089), and a nonsignificant number of infants (8/16) showed the effect. In Study 4, however, the effect of trial type was significant, the effect size was larger (partial η^2 = .341), and a significant number of infants (12/16) showed the effect. To follow up these separate analyses, we carried out a direct between-study comparison. We combined the search duration data from Studies 4 and 5 and conducted a 2 × 2 × 2 mixed ANOVA, with lexical category (adjectives in Study 5, proper names in Study 4) and trial order (two words from same lexical category first, two words from same lexical category second) as between-subjects factors, and trial type (two words from same lexical category, two words from different lexical categories) as a within-subjects factor. In this follow-up analysis, there was a significant effect of trial type, $F(1, 28) = 6.52$, $p = .016$, partial η^2 = .189, but the expected interaction between lexical category and trial type was not significant, $F(1, 28) = 0.46$, $p = .50$, partial η^2 = .016.

Close inspection of the results from Studies 4 and 5 reveals a way to reconcile the apparent conflict between the nonsignificant trial type effect in the analysis of Study 5 and the nonsignificant interaction between lexical category and trial type in the cross-study analysis. In Study 5, although the trial type effect was nonsignificant, the difference between trial means was in the predicted direction, indicating that infants had a weak tendency to expect two objects upon hearing two adjectives. However, the trial type effect was more than three times *larger* when the study involved proper names (partial η^2 = .341) than when it involved adjectives (partial η^2 = .089). The stronger effect in Study 4 (with proper names) and the weaker effect in Study 5 (with adjectives) thus lend support to two claims. First, infants *did* distinguish between the use of proper names and adjectives in our task. This discovery suggests that the effect uncovered in Study 4 reflects infants' knowledge of the semantic distinction between proper names and count nouns, along with adherence to the principle of contrast. Second, infants also had a weaker pragmatic tendency to infer that a speaker who used two words from the same lexical category (either two proper names or two adjectives) intended the words to apply to two different objects.

General Discussion

The results of these five studies provide the first evidence in the literature that infants' emerging understanding of distinct lexical-category-to-meaning links guides them in the fundamental task of object individuation. The findings make three main contributions. First, they reveal a previously unidentified source of information available to drive individuation in infancy. Prior work has established that by 12 months (or earlier), infants can use information about both spatiotemporal differences and property–kind differences between objects to guide individuation (Bonatti, Frot, Zangl, & Mehler, 2002; Wilcox & Baillargeon, 1998; Wilcox & Schweinle, 2002; Xu & Baker, 2005; Xu & Carey, 1996; Xu, Carey, & Quint, 2004). Previous research has also identified a role for language. For example, 10-month-old infants often fail to use property–kind differences between objects to individuate them in a complex task, but they may succeed when the objects receive distinct count noun labels (Xu, 2002). Moreover, Xu et al. (2005) found that 12-month-olds who heard two different novel count nouns were more likely to infer the presence of two objects hidden inside an opaque box than those who heard one count noun repeated. The current work reveals that, in addition to these cues, 20-month-old infants can use lexical category information to guide object individuation. Specifically, they can exploit this information to disambiguate an event that involves two sequential appearances of an object.

It would be interesting to push this line of research further by removing the presence of the object(s) from the task. The results from the current task do not allow us to determine whether infants needed to see the object appearances in order to use the lexical category information appropriately in drawing inferences about the number of objects hidden in the box. Eliminating the objects from the task would enable a test of whether infants can use lexical category information *alone* to set up a mental model of the number of objects present in an ambiguous scene. Moreover, if we take the presentation of one or more objects to be a defining property of an object individuation task, infants' success on this modified task would suggest that they have the ability to use lexical category cues to predict the number of objects present in the ambiguous situation, without individuating objects per se. In future work, it would be interesting to examine infants' use of lexical category information in this type of modified task (see Dewar & Xu, 2007; Xu et al., 2005).

Second, the results provide the first evidence from a manual search task that by 20 months, infants have acquired knowledge of appropriate semantic distinctions among three word categories (proper names, count nouns, and adjectives). The results are consistent with a body of research reporting knowledge of several distinct lexical-category-to-meaning links in children younger than 2 years (e.g., Booth & Waxman, 2003; Echols & Marti, 2004; Waxman, 1999; Waxman & Booth, 2001; Waxman & Markow, 1998). More specifically, the current results corroborate recent findings from a study that used a categorization task involving word extension to investigate infants' knowledge of the distinct semantics of proper names and count nouns (Bélanger & Hall, 2006). In that study, as in the current one, 20-month-olds but not 16-month-olds distinguished appropriately between these two lexical categories. The agreement between the results of the two studies provides evidence that lexical category knowledge is available to guide infants in a manual search task as soon as it can be used to direct them in word extension. The convergence between the search and the word extension results also bolsters the claim that knowledge of lexical-category-to-meaning links for proper names and count nouns first arises between 16 and 20 months.

Finally, the findings offer new evidence that infants use the principle of contrast in the acquisition of novel words. The effects we observed among 20-month-olds cannot be due solely to their knowledge of the semantics of different lexical categories. Infants must also have an understanding of circumstances under which the use of two novel words implies reference to two different objects, such as when the two words would violate the contrast principle (Clark, 1993) if they referred to just one object. In our task, two novel proper names for one object would violate the contrast principle because both words would have the same meaning (i.e., the words would both refer to the *same* individual object). Two novel adjectives for one object, however, would not necessarily violate the contrast principle because they could be taken to have different meanings (i.e., the words could refer to *different* object properties). The combination of a count noun followed by a proper name or a count noun followed by an adjective for one object would not violate contrast because the two terms would have different meanings (i.e., one would refer to an object category and the other to an individual object or an object property). It is possible that infants represent the contrast principle as a semantic constraint on the possible meanings that can be assigned to novel words, but it is also possible that it is a pragmatic assumption about a speaker's intended use of words within a discourse context. One piece of evidence favoring the pragmatic account is that many objects do have more than one proper name (e.g., one man may be called "William," "Bill," or "Mr. Smith" by different people and/or in different settings); however, it is unclear whether 20-month-olds possess this knowledge (for further discussion of pragmatics and the contrast principle, see Bloom, 2000; Clark, 1993; Diesendruck, 2005; Diesendruck & Markson, 2001).

The results from Study 5 also suggest that infants have an additional, weaker expectation that a speaker who uses two novel words from a single lexical category intends to refer to two different objects, even when the words would not necessarily violate the contrast principle if applied to one object (as in the case of two novel adjectives). Further research manipulating the conditions under which infants are exposed to multiple novel words for objects should help clarify the nature of their assumptions about the possible semantic relations that can hold among them.

References

Bélanger, J., & Hall, D. G. (2006). Learning proper names and count nouns: Evidence from 16- and 20-month-olds. *Journal of Cognition and Development*, 7, 45–72.

Bélanger, J., & Hall, D. G. (2008). *The origins of lexical-category-to-meaning links: The case of count nouns and proper names. Manuscript in preparation.* University of British Columbia, Vancouver, Canada.

Bloom, P. (2000). *How children learn the meanings of words.* Cambridge, MA: MIT Press.

Bonatti, L., Frot, E., Zangl, R., & Mehler, J. (2002). The human first hypothesis: Identification of conspecifics and individuation of objects in the young infant. *Cognitive Psychology, 44*, 388–426.

Booth, A., & Waxman, S.R. (2003). Mapping words to the world in infancy: Infants' expectations for count nouns and adjectives. *Journal of Cognition and Development, 4*, 357–381.

Clark, E. (1993). *The lexicon in acquisition*. Cambridge, UK: Cambridge University Press.

Dewar, K., & Xu, F. (2007). Do 9-month-old infants expect distinct words to refer to kinds? *Developmental Psychology, 43*, 1227–1238.

Diesendruck, G. (2005). The principles of conventionality and contrast in word learning: An empirical examination. *Developmental Psychology, 41*, 451–463.

Diesendruck, G., Hall, D.G., & Graham, S.A. (2006). Children's use of syntactic and pragmatic knowledge in the interpretation of novel adjectives. *Child Development, 77*, 16–30.

Diesendruck, G., & Markson, L. (2001). Children's avoidance of lexical overlap: A pragmatic account. *Developmental Psychology, 37*, 630–641.

Echols, C., & Marti, C. N. (2004). The identification of words and their meanings: From perceptual biases to language-specific cues. In D. G. Hall & S. Waxman (Eds.), *Weaving a lexicon* (pp. 41–78). Cambridge, MA: MIT Press.

Feigenson, L., & Carey, S. (2003). Tracking individuals via object files: Evidence from infants' manual search. *Developmental Science, 6*, 568–584.

Fisher, C., Klingler, S. L., & Song, H. (2006). What does syntax say about space? 2-year-olds use sentence structure to learn new prepositions. *Cognition, 101*, B19–B29.

Hall, D. G., & Graham, S. A. (1999). Lexical form class information guides word-to-object mapping in preschoolers. *Child Development, 70*, 78–91.

Hall, D. G., Lee, S. C., & Bélanger, J. (2001). Young children's use of syntactic cues to learn proper names and count nouns. *Developmental Psychology, 37*, 298–307.

Liittschwager, J., & Markman, E. (1993, March). *Young children's understanding of proper versus common nouns.* Poster presented at the biennial meetings of the Society for Research in Child Development, New Orleans, LA.

Markman, E. M. (1989). *Categorization and naming in children*. Cambridge, MA: MIT Press.

Sorrentino, C. (2001). Individuation, identity, and proper names in cognitive development. *Developmental Science, 4*, 399–407.

Sorrentino, C., & Halberda, J. (2001). *Do multiple proper names indicate multiple individuals? Evidence from children and adults.* Poster presented at the biennial meeting of the Society for Research in Child Development, Minneapolis, MN.

Van de Walle, G., Carey, S., & Prevor, M. (2000). Bases for object individuation in infancy: Evidence from manual search. *Journal of Cognition and Development, 1*, 249–280.

Waxman, S. R. (1999). Specifying the scope of 13-month-olds' expectations for novel words. *Cognition, 70*, B35–B50.

Waxman, S. R., & Booth, A. E. (2001). Seeing pink elephants: Fourteen-month-olds' interpretations of novel nouns and adjectives. *Cognitive Psychology, 43*, 217–242

Waxman, S. R., & Markow, D. B. (1998). Object properties and object kinds: Twenty-one-month-old infants' extension of novel adjectives. *Child Development, 69*, 1313–1329.

Wilcox, T., & Baillargeon, R. (1998). Object individuation in infancy: The use of featural information in reasoning about occlusion events. *Cognitive Psychology, 37*, 97–155.

Wilcox, T., & Schweinle, A. (2002). Object individuation and event mapping: Developmental changes in infants' use of featural information. *Developmental Science, 5*, 132–150.

Xu, F. (2002). The role of language in acquiring object kind concepts in infancy. *Cognition, 85*, 223–250.

Xu, F., & Baker, A. (2005). Object individuation in 10-month-old infants using a simplified manual search method. *Journal of Cognition and Development, 6*, 307–323.

Xu, F., & Carey, S. (1996). Infants' metaphysics: The case of numerical identity. *Cognitive Psychology, 30*, 111–153.

Xu, F., Carey, S., & Quint, N. (2004). The emergence of kind-based object individuation in infancy. *Cognitive Psychology, 49*, 155–190.

Xu, F., Cote, M., & Baker, A. (2005). Labeling guides object individuation in 12-month-old infants. *Psychological Science, 16*, 372–377.

Child Development, September/October 2008, Volume 79, Number 5, Pages 1444 – 1462

Development of Emotional and Behavioral Regulation in Children Born Extremely Preterm and Very Preterm: Biological and Social Influences

Caron A. C. Clark
University of Canterbury

Lianne J. Woodward
University of Canterbury and Van der Veer Institute for Parkinson's and Brain Research

L. John Horwood and Stephanie Moor
University of Otago

This study describes the development of emotional and behavioral regulation in a regional cohort of children born extremely preterm (<28 weeks gestational age, n = 39), very preterm (<34 weeks gestational age, n = 56), and full term (n = 103). At 2 and 4 years, children born at younger gestational ages demonstrated poorer self-regulation across multiple contexts spanning observed interactions, formal cognitive testing, and parental report of child behavior at home. Among children born preterm, the 2 strongest predictors of impairments in self-regulation were the presence of moderate-to-severe cerebral white matter abnormalities on neonatal magnetic resonance and a less sensitive parenting style when children were aged 2 years. Findings support the importance of early neurological development and parenting for developing regulation in children born very preterm.

As more sophisticated medical intensive care has rescued ever smaller babies, increasing concerns about the longer term social and emotional consequences for these children have been raised. Findings suggest that, by middle childhood, children born very preterm are almost three times more likely to meet *Diagnostic and Statistical Manual of Mental Disorders (DSM)* criteria for attention-deficit hyperactivity disorder and are at elevated risk for depression, anxiety disorders, and behavioral problems compared to their full-term peers (Bhutta, Cleves, Casey, Cradock, & Anand, 2002; Botting, Powls, Cooke, & Marlow, 1997; Breslau et al., 1996; Farooqi, Hagglof, Sedin, Gothefors, & Serenius, 2007; Horwood, Mogridge, & Darlow, 1998). Relatedly, very preterm children have also been found to have higher rates of social withdrawal and peer victimization (Breslau et al., 1996; Gardner et al., 2004; Nadeau, Boivin, Tessier, Lefebvre, & Robaey, 2001; Nadeau, Tessier, Boivin, Lefebvre, & Robaey, 2003).

Despite initial optimism that these emotional and behavioral difficulties would improve with age, growing evidence now suggests that these problems often persist into adolescence and, to some extent, early adulthood (Cooke, 2004; Gardner et al., 2004; Hack et al., 2002; Indredavik, Vik, Heyerdahl, Kulseng, & Brubakk, 2005). For example, 15- to 16-year-old teenagers born extremely preterm (≤ 28 weeks gestation) have been found to report 2 to 3 times more emotional problems and to have lower rates of normative peer risk taking and social engagement than teenagers born full term (Gardner et al., 2004). Parents and teachers also rated these teens as showing higher rates of emotional problems, hyperactivity, and peer relationship problems, with these difficulties impacting on home life, learning, friendships, and leisure. These findings highlight the presence of persistent mental health and social impairments among children and adolescents born very preterm. Furthermore, there is evidence to suggest that risks of later problems are greatest among those born at the lower limits of viability, that is, extremely preterm (< 28 weeks) or extremely low birth weight (< 1,000 g; Hack et al., 1994; Klebanov, Brooks-Gunn, & McCormick, 1994).

This research was funded by grants from the Neurological Foundation of New Zealand, Lottery Grants Board, Canterbury Medical Research Foundation, and the Health Research Council of New Zealand. Special thanks to Associate Professor Terrie Inder (pediatric neurologist) and Dr. Scott Wells (radiologist) for MR imaging analysis, others in the research team (Dr. Patricia Champion, Dr. Nicola Austin, Jacqueline Knight, Kelly Hood, Michelle VanDyk, Michelle Davey. and Carole Spencer), and most importantly to the study families for their willingness to share their children's lives with us.

Correspondence concerning this article should be addressed to Caron A. C. Clark, Canterbury Child Development Research Group, Department of Psychology, University of Canterbury, Private Bag 4800, Christchurch, New Zealand. Electronic mail may be sent to cac65@student.canterbury.ac.nz.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7905-0015

Although clear linkages have been demonstrated between preterm birth and children's later psychosocial risks during childhood and adolescence, much less is known about the early developmental precursors of these later difficulties, the manner in which such difficulties may develop and manifest over early childhood, or the clinical and socioenvironmental factors that place very preterm children at later risk. Such information is essential for the identification of at-risk infants and the development of effective early intervention strategies to assist these children and their families. Accordingly, the goal of this study was to address these important research gaps using prospective longitudinal data from a regional cohort of very preterm infants (a) to examine the early development of emotional and behavioral regulation of very preterm children and (b) to examine, within the preterm group, clinical and socioenvironmental predictors of early regulatory delay and/or impairment. A review of existing evidence relating to the importance of self-regulation for psychosocial development, as well as a consideration of the factors that may place children born preterm at elevated risk of regulatory problems, is provided below.

Self-Regulation and Homeostasis in the Preterm Infant

Self-regulation refers to the ability to modulate emotion, self-soothe, delay gratification, and tolerate change in the environment (DeGangi, 2000; Kopp, 1982). Temperament theorists also emphasize key biological processes such as attention, approach, avoidance, and inhibition in the modulation of arousal and reactivity (Goldsmith et al., 1987). The toddler and early preschool years represent an important period for the emergence of self-regulation (Kopp, 1982). Internal homeostasis allows children to maintain optimal arousal, attend to the external environment, explore, and learn. Over this period, drawing on internal working models gained from dyadic attachment relationships and biological maturational processes, children gradually become more efficient at postponing gratification, soothing and calming themselves to prevent emotional lability, and sustaining and flexibly manipulating their attention. Such abilities are integral to the development of advanced social, emotional, and cognitive function (Eisenberg et al., 2000; Lengua, 2003).

Given the importance of self-regulation for children's emerging emotional, behavioral, and social development, it has increasingly been recognized as a potentially important marker of later psychosocial risk (Cole, Michel, & Teti, 1994; Lawson & Ruff, 2004). Certainly, good theoretical support exists for the role of early emotional dysregulation in anxious-depressive disorders (Calkins, 1994; Dahl, 2003). Also, although learning difficulties and attentional disorders are commonly only diagnosed following school entry, children who develop such difficulties have been found to show signs of poor attentional regulation as preschoolers (Lawson & Ruff, 2004). Given these findings, self-regulation is now increasingly recognized as an important developmental task of early childhood.

There are several reasons to suspect that children born very preterm may be vulnerable to regulatory difficulties in early childhood. In particular, clinical and observational studies suggest that very young preterm infants are more excitable and irritable, have more difficulty regulating their arousal and motor responses, and have more difficulty directing their attention to salient stimuli (Brown, Doyle, Bear, & Inder, 2006; Eckerman, Oehler, Medvin, & Hannan, 1994; Hughes, Shults, McGrath, & Medoff-Cooper, 2002; Mouradian, Als, & Coster, 2000). Preterm infants characterized by higher levels of medical risk have also been found to have more negative facial expressions, disorganized movement, and poorer attentional control when overstimulated compared to their lower risk, preterm peers (Eckerman et al., 1994). In another study, parents of very preterm infants reported that their infants had less regular sleeping, feeding, and elimination patterns; were less intense in their responses; and were more withdrawing from new situations, stimuli, or places than full-term infants (Hughes et al., 2002). Although some improvement was evident on measures of activity and regularity by age 6 months, preterm infants remained less adaptable. Consistent with these parent report studies, several observational studies have shown that trained independent observers tend to rate very preterm infants as showing fewer positive and more negative expressions and as exhibiting more difficulties transitioning across tasks, compared to term-born infants (Landry, Garner, & Swank, 1996; Segal et al., 1995). Furthermore, preterm infants who have sustained brain insult such as intraventricular hemorrhage have also been found to display less sociability and less behavioral approach, suggestive of dampened behavioral reactivity (Garcia Coll et al., 1988).

Taken together, despite differences in methodological approaches (neurobehavioral examination, temperament scales, and direct observation), there is now converging evidence that, during the 1st year of life, infants born very preterm experience difficulties in the regulation of arousal and emotional responses and in flexibly allocating and sustaining their attention. Furthermore, as has been shown with later cognitive and behavioral outcomes, there is evidence of a dose–response relationship, with risks of later difficulties increasing with decreasing gestational age (Duffy, Als, & McAnulty, 1990; Mouradian et al., 2000). However, the extent to which these regulatory difficulties persist or change over the toddler and preschool years has yet to be established.

Clinical and Sociofamilial Factors Associated With Later Regulatory Abilities in Children Born Very Preterm

A second issue raised by these childhood findings concerns the clinical and socioenvironmental risk factors that place very preterm children at elevated risk of early and persistent regulatory difficulties. To date, few studies have examined the role of perinatal factors such as the degree of immaturity, infant illness severity, or the extent of early cerebral neuroanatomical abnormalities in accounting for individual differences in regulation and behavioral adjustment. It is possible that the artificial environment of the neonatal intensive care unit, accompanied by the pain and stress of these medical procedures, may have ongoing consequences for children's homeostatic regulation of arousal (Graham, Heim, Goodman, Miller, & Nemeroff, 1999; Porges, 1996). For example, infants who are exposed to more painful medical procedures have been found to have higher levels of cortisol at follow-up points during the first 2 years of life, suggesting that these early medical experiences may permanently alter the set point of arousal in these infants (Grunau, Weinberg, & Whitfield, 2004).

Accompanying these medical experiences, neonatal magnetic resonance (MR) imaging studies indicate that the majority of infants born preterm have white matter abnormalities, including signal abnormalities, loss of volume, cystic abnormalities, enlarged ventricles, thinning of the corpus callosum, and decreased myelination (Inder, Warfield, Wang, Huppi, & Volpe, 2005). Although some studies have demonstrated a relationship between neurological abnormalities and neurobehavior in these infants at term age (Brown et al., 2006; Mouradian et al., 2000), no studies to date have examined the relationship between MR findings and emotional and behavioral outcomes in preschool-aged children born very preterm.

Even less consideration has been given to the contribution of social contextual factors such as the availability of socioeconomic resources, maternal psychological well-being, and parenting, despite clear evidence demonstrating the importance of these factors for children's behavioral development (Colman, Hardy, Albert, Raffaelli, & Crocket, 2006; Shaw, Keenan, Vondra, Delliquadri, & Giovannelli, 1997; Supplee, Shaw, Hailstones, & Hartman, 2004). This is important, given some suggestion that children born very preterm who are at high medical risk may benefit from parenting that is sensitive and nonintrusive in timing and nature (Forcada-Guex, Pierrehumbert, Borghini, Moessinger, & Muller-Nix, 2006; Landry et al., 1996). Other contextual factors that may be influential include lower socioeconomic resources, lower maternal education, and maternal depression (Murray & Cooper, 1997; O'Connor, Heron, Glover, & Alspac Study, 2002; Silk, Shaw, Skuban, Oland, & Kovacs, 2006). Such risk factors have been shown to be more common among parents of preterm infants (Kersting et al., 2004; Miles, Holditch-Davis, Schwartz, & Scher, 2007). Examination of both clinical and sociofamilial factors will be important in developing a better understanding of the processes shaping the developmental pathways toward regulatory competence in children born preterm.

Two obstacles to a better understanding of the development of self-regulation in this high-risk population are the paucity of reliable and valid measures of self-regulation in preschool children and the tendency for many studies to rely solely on parent report measures of infant temperament and behavior (e.g., Delobel-Ayoub et al., 2006; McCormick, Workman-Daniels, & Brooks-Gunn, 1996). Porges (1996) has recommended that the measurement of self-regulatory capacities be based on behaviors such as the quality, quantity, and appropriateness of body movements and coordination of emotional responses. Thus, behavioral observation in a demanding task environment may provide a valuable complement to parent report measures of self-regulation. In addition, others have stressed the importance of context and the goal at hand in assessing a child's regulatory competence (Cole et al., 1994; Thompson, 1994). For example, although particular strategies for the regulation of emotion may be appropriate at one age or in one context, they may become less adaptive or acceptable as children age or as they enter more challenging social situations. Thus, effective self-regulation is multidimensional, influenced by contextual demands, and reliant on a range of processes. These processes span arousal and attentional flexibility, the expression of affect, and the ability to transition fluently across situations.

Drawing on these conceptual and measurement recommendations, we adopted a multimethod measurement approach in order to examine the development of preschool children's emotional and behavioral regulation at ages 2 and 4 years in a cohort of children born preterm and full term. Given previous evidence showing that children born extremely preterm may be especially vulnerable to later developmental challenges, the effect of gestational age on later regulatory outcomes was also considered. We hypothesized that, relative to their full-term peers, children born preterm would show impaired emotional and behavioral regulation across all assessment contexts, with those born extremely preterm being most impaired. The specific aims of the study were as follows.

1. To describe the emotional and behavioral regulation of children born extremely preterm, very preterm, and full term at ages 2 and 4 years (corrected for gestational age at birth). Repeated measures of regulation across the two time points included observed child behavior during structured parent – child interaction and during cognitive testing, in addition to parent report.
2. To describe the development of self-regulation over the period from 2 to 4 years in children born very (< 33 and > 28 weeks) and extremely (≤ 28 weeks) preterm.
3. To examine, within the preterm sample, the extent to which impairment in the development of self-regulation between the ages of 2 and 4 years could be explained by a range of risk factors spanning infant clinical and sociofamilial factors. These included intrauterine growth restriction (IUGR, defined as a birth weight below the 10th percentile for gestational age), chronic lung disease (defined as oxygen use at 36 weeks), patent ductus arteriosis (PDA, a circulatory condition arising as a consequence of an immature heart), administration of antenatal steroids, necrotizing enterocolitis (NEC, a condition of the intestine), white matter abnormalities as measured by MR, maternal education, family socioeconomic status (SES), maternal mental health, experience of early intervention, and parenting style.

Method

Participants

The study included three groups of children who were recruited as part of a prospective longitudinal study aimed at examining the developmental and neuropsychological outcomes associated with prematurity. These study groups are described below.

Children Born Very Preterm and Extremely Preterm

The first two groups were drawn from a regional cohort of 102 children born very preterm (≤ 33 weeks) and/or very low birth weight (< 1,500 g) who were admitted consecutively to a Level III neonatal intensive care unit at Christchurch Women's Hospital, New Zealand from November 1998 to February 2001. Families were invited to participate in the study by a research nurse based at the hospital. Exclusion criteria included infant congenital abnormalities and non-English-speaking parents. Ninety-two percent of all eligible infants were recruited. These infants had a mean birth weight of 1,069 g (range = 385 – 1,790 g) and a mean gestational age of 28 weeks (range = 23 – 33 weeks). All children born preterm were assessed throughout the perinatal period, at term equivalent, and at 2 and 4 years (corrected for gestational age at birth). Excluding deaths (n = 3), sample retention to age 2 years was 97%, with 2 children untraced and 1 overseas relocation. An additional child was unable to be assessed due to severe visual impairment. At age 4 years, sample retention was 99%. Although one family refused to participate at this follow-up point, the 3 children who were not seen at age 2 years were successfully retraced at age 4 years. Because prior research has suggested that gestational age is a robust predictor of developmental outcome, all children born preterm were divided into two groups: (a) those born between 28 and 33 weeks gestation (n = 59) and (b) those born prior to 28 weeks gestation (n = 39). Sample stratification was based on gestational age rather than birth weight to avoid the confounding effects of IUGR. These gestational cut points are widely used to define the extent of immaturity at birth and have good predictive validity (Anderson, Doyle, & the Victorian Infant Collaborative Study, 2003; Doyle, Casalaz, & the Victorian Infant Collaborative Study Group, 2001; Foster-Cohen, Champion, Edgin, & Woodward, 2007; Horwood et al., 1998; Marlow, 2004; Nadeau et al., 2001). In the interests of consistency, these groups of children are collectively referred to as children born preterm.

Children Born Full Term

In addition to the preterm groups, a group of 103 children born full term (> 37 weeks gestation at birth) was recruited at age 2 years. These children were identified from the hospital birth record database by selecting, in an alternative manner, a child of the same gender born two deliveries prior to, or following, the birth of each preterm child. These children had a mean gestational age of 40 weeks (range = 38 – 41 weeks) and a mean birth weight of 3,592 g (range = 2,700–4,630 g). Consistent with the preterm groups, children with congenital abnormalities and those from non-English-speaking families were excluded. Of the families identified, 62% participated in the study. Of those not recruited, 47% were untraced, 12.5% had relocated overseas, 12.5% refused, and 28% initially agreed to participate but were unable to attend their 2-year assessment within the specified time frame (±2 weeks) due to illness or family circumstances. No significant between-group differences were found between nonrecruited and recruited term-born children on measures of gestational age, birth weight, family SES, or the proportion of Maori families. (The Maori population are the indigenous people of New Zealand. Although they make up about 12% of the general population, only 10% of all Maori live in the South Island; Statistics New Zealand, 2001.) At age 4 years, 96% of this sample was reassessed, with 2 children unable to be traced and 2 families declining to participate.

For children participating in the 2- and 4-year follow-up assessments, no significant differences were found across the three groups on measures of child gender ($p = .92$), ethnic distribution ($p = .64$), or maternal age at birth ($p = .89$). However, children born very preterm or extremely preterm were more likely to come from families of lower SES ($p < .001$), and their mothers were more likely to have left school before the age of 16 years ($p < .01$).

Measures

Figure 1 describes the data collection points and the key measures obtained for this study. As illustrated, a series of common measures assessing aspects of children's emotional, behavioral, and attentional regulation were administered at the ages of 2 and 4 years. These measures were selected based on existing research and theory and included (a) observational ratings of children's behavior during a structured parent–child interaction task, (b) independent tester ratings of child behavior during cognitive assessments, and (c) an adaptation of the parent-reported Emotion Regulation Checklist (Shields & Cicchetti, 1997). A detailed description of each of these study measures is given below.

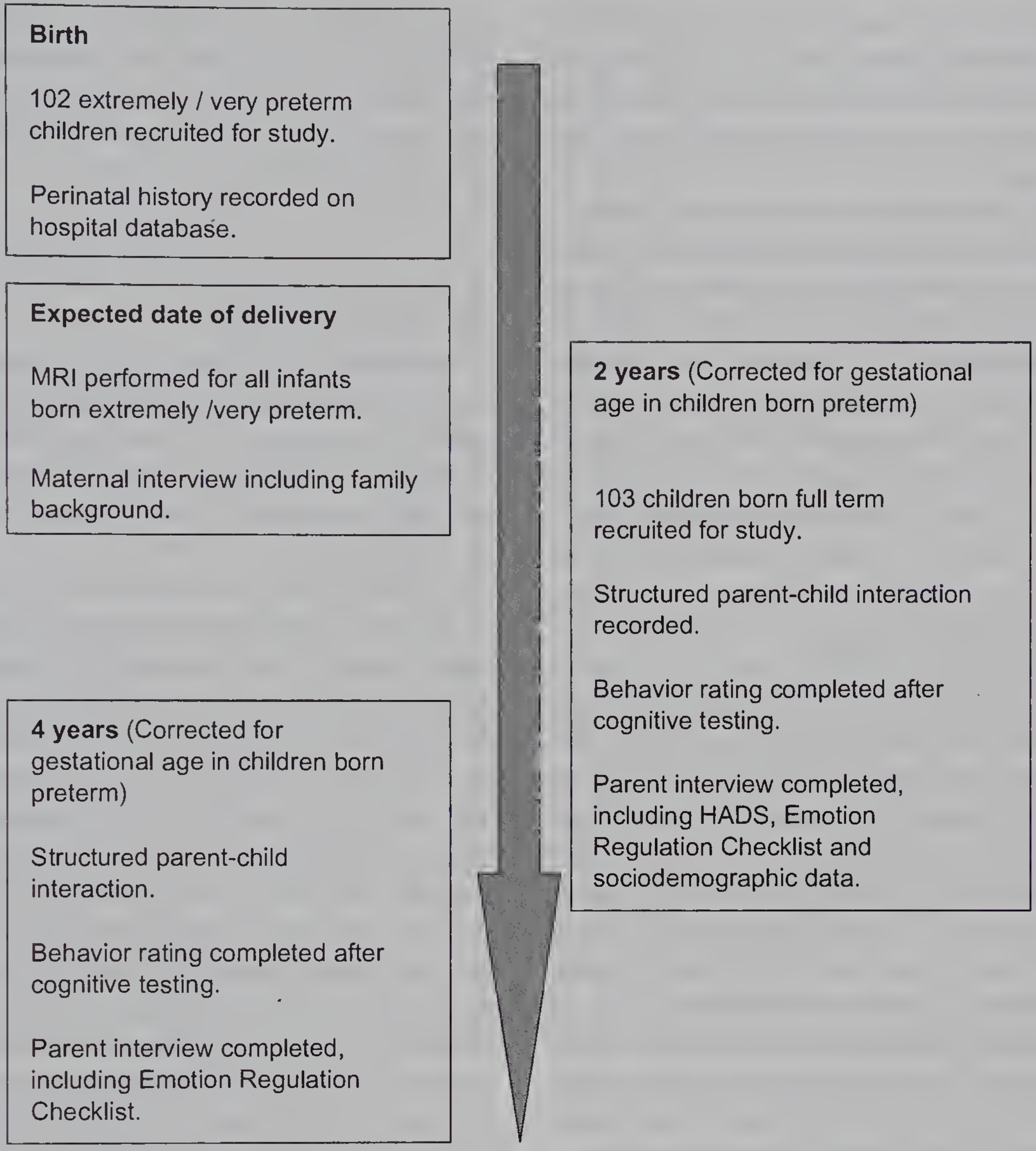
Emotional and Behavioral Regulation Measures

Observed child behavior during parent–child interaction. At ages 2 and 4 years, at the beginning of each child's developmental assessment, a structured parent–child play interaction took place. This was videotaped for subsequent coding and analysis. This activity was described to parents as an opportunity to see how their child learned to solve problems with the support of someone familiar. Before beginning the activity, each problem-solving task was explained to the parent and they were shown how to complete the task while another researcher distracted the child with a toy. Parents were instructed to introduce and complete each task one at a time and to allow their child a few minutes to explore each task by themselves before offering assistance. Following this time, parents were told that they could provide whatever assistance they deemed necessary to help their child complete the task. Once the parent indicated that they understood the procedure, the researchers left the room and video recording began.

Three different age-appropriate problem-solving tasks were used at each assessment point. Tasks were selected to be challenging for the child and were ordered, based on piloting data, to increase in difficulty. The tasks used at both ages were similar in nature but differed in complexity to offer appropriate developmental challenge. At age 2 years, children and their parents completed three activities. The first was a color-matching puzzle, which involved matching 12 colored disks to their corresponding color on a puzzle board. The second task required the child to move a sliding lid on the top of a two-holed (triangle and rectangle) posting box so as to place each shape in the corresponding posting hole. The third activity involved a more complex task using a clear, rectangular perspex box (Matas, Arend, & Sroufe, 1978). A small smiley-faced stamp was placed within the box, and the child had to place a weight on a wooden lever to lift the stamp and retrieve it.

At age 4 years, children completed four tasks. The first task was a simple, 10-piece jigsaw puzzle of a teddy bear. The second activity was a magnetic-letter board, requiring children to copy the word "flower" using magnetic letters. All letters of the alphabet were available, and children had to select those of the same color to match on the board. The third activity required children to copy a castle model using large Lego blocks, and the fourth was a bead-threading task in which children were required to reproduce a pattern of threaded beads of different colors and shapes. The activities were hidden with cloth covers so that parents could introduce each activity one at a time.

A range of child and parent behaviors were coded, based on these videotaped parent–child interactions. Raters coded child behaviors first and then rewound the tape to code parent behaviors. For each problem-solving task, four child behaviors were rated using



Birth

102 extremely / very preterm children recruited for study.

Perinatal history recorded on hospital database.

Expected date of delivery

MRI performed for all infants born extremely /very preterm.

Maternal interview including family background.

2 years (Corrected for gestational age in children born preterm)

103 children born full term recruited for study.

Structured parent-child interaction recorded.

Behavior rating completed after cognitive testing.

Parent interview completed, including HADS, Emotion Regulation Checklist and sociodemographic data.

4 years (Corrected for gestational age in children born preterm)

Structured parent-child interaction.

Behavior rating completed after cognitive testing.

Parent interview completed, including Emotion Regulation Checklist.

Figure 1. Overview of study design.

5-point Likert scales. These behaviors were based on previously used coding schemes and included positive affect, negative affect, persistence/sustained attention, and the quality of transition across tasks (Belsky, Jaffee, Sligo, Woodward, & Silva, 2005; Chase-Lansdale, Brooks-Gunn, & Zamsky, 1994; National Institute of Child Health and Human Development Early Child Care Research Network, 1999). *Positive affect* was coded on the basis of the intensity and duration of positive facial expressions (smiles and laughter), exclamations, and voice tone. *Negative affect* assessed the frequency and intensity of frowns, pouting, whining, crying, and signs of anger or frustration. *Persistence* was based on the level of attention and motivation the child showed while completing the task. Children who actively avoided completing a task were allocated low ratings, whereas those who worked persistently to complete a task were allocated the highest ratings. *Quality of a child's transition* between tasks was coded based on how smoothly a child moved from the completion of one problem-solving task to another. Children who refused to engage in new tasks or returned persistently to a previous task were rated lower on this scale. An overall mean rating for each child behavior was computed by averaging scores across the three tasks completed at age 2 years and the four tasks completed at age 4 years.

Two raters independently coded tapes using pencil-and-paper methods at age 2 years. Three raters independently coded tapes for the 4-year interactions. All observers were trained by the same clinical psychologist and were blind to infant perinatal history. Interrater agreement was assessed for 15% of tapes at 2 years and 20% at 4 years. Pearson product moment correlations between the two raters at age 2 years ranged from .68 to .90. Intraclass correlations were used at age 4 years because there were three raters at this age. Correlations for absolute agreement ranged from .83 to .90 for the four codes at age 4 years.

Complete video data were not available for 8 children at 2 years and 3 children at 4 years due to problems with recording quality or video recorder malfunction. These children were therefore excluded from analyses at the respective ages.

Tester ratings of child behavior during cognitive assessment. Children completed the Bayley Scales of Infant Development—Revised (Bayley, 1993) at age 2 years. At age 4 years, children completed an extensive battery of executive function tasks, as well as a four-subtest version of the Wechsler Preschool and Primary Scales of Intelligence—Revised (Wechsler, 1989). Cognitive and health outcomes of this cohort are reported elsewhere (Edgin et al., 2007; Foster-Cohen et al., 2007; Gray, Woodward, Spencer, Inder, & Austin, 2006; Woodward, Anderson, Austin, Howard, & Inder, 2006; Woodward, Edgin, Thompson, & Inder, 2005). Following cognitive testing at each age, two examiners completed consensus ratings of children's behavior based on an abbreviated version of the Behavior Rating Scale of the Bayley Scales of Infant Development at age 2 years. This scale included items assessing child positive affect, negative affect, initiative, attention, persistence, enthusiasm toward tasks, orientation to examiner, social engagement, and cooperation. Each item was rated on a scale ranging from 1 (*less observable*) to 5 (*more observable*). These measures were chosen based on the previous literature suggesting that children born preterm may have difficulties in these areas and because they were deemed to be developmentally relevant for both 2- and 4-year-olds. Examiners were blind to gestational age and clinical history but were not always able to remain blind to group status. A composite measure of tester-rated regulation was computed from the mean of these items. Cronbach's alpha coefficients for this nine item scale were .93 at age 2 and .90 at age 4 years.

Parent-rated self-regulation. During each assessment, parents completed a structured interview with a trained research nurse. A modified version of the Emotion Regulation Checklist (Shields & Cicchetti, 1997) was included in this interview. Scale items were reviewed by both a clinical and a developmental psychologist and were reworded to be appropriate for young children in New Zealand. The final checklist consisted of 24 items assessing the management and lability of emotions and tolerance to upset. Statements such as "is prone to disruptive outbursts of energy or excitement" and "responds positively to neutral or friendly attempts to interact with peers" were rated on a 3-point scale ranging from 1 (*not true*) to 3 (*certainly true*). Some of these items were reverse coded in order to provide a composite measure of emotion regulation, with higher scores reflecting a greater ability to manage one's emotion. For analysis purposes, scores were rescaled so that they were on the same metric (1–5) as observational measures. Cronbach's alpha coefficients for this scale were .69 at age 2 and .64 at age 4 years.

Infant Clinical, Neuroanatomical, and Family Background Measures

To examine the extent to which the development of self-regulation across the three study groups over the period from age 2 to 4 years was influenced by children's clinical history and sociofamilial experiences after birth, the following measures were included in this analysis.

Infant clinical and neurological characteristics. Measures of infant health and well-being included gender, birth weight, IUGR, twin birth status, chronic lung disease, antenatal steroid exposure, NEC, and PDA. In addition, at the time of their expected date of delivery, all infants born very preterm underwent MR imaging using a 1.5 T General Electric Signa System Scanner (GE Medical Systems, Milwaukee, WI). Infants were calmed, wrapped, and positioned in the scanner in a vacuum fixation pillow, which limited movement and allowed for scanning with no sedation. Two imaging modes were applied. These included (a) a three-dimensional Fourier transform spoiled gradient recalled sequence (1.5 coronal slices, flip angle 45°, repetition time 35 ms, echo time 5 ms, field of view 18 cm, matrix 256 × 256) and (b) a double echo (proton density and T2-weighted) spin echo sequence (DE; 3 mm axial slices, repetition time 3,000 ms, echo times 36 and 162 ms, field of view 18 cm, matrix 256 × 256, interleaved acquisition). The voxel volumes were 0.7 × 0.7 × 1.5 mm for the spoiled gradient recalled sequence acquisition and 0.7 × 0.7 × 3 mm for the spin echo acquisition. Full details of the scanning protocol have been reported in previous articles (Inder, Anderson, Spencer, Wells, & Volpe, 2003; Woodward et al., 2005).

MR scans were analyzed on five 3-point scales assessing (a) the extent of white matter signal abnormality, (b) periventricular white matter volume loss, (c) the presence of cysts, (d) ventricular dilation, and (e) thinning of the corpus callosum and reduced myelination. A pediatric radiologist, who was blind to the medical history of the infants, scored each scan. Each scan was then independently reviewed by another pediatric neurologist, with discrepant cases given a consensus rating. Interrater agreement was 95%. One child in the preterm group was unable to be scored. Based on their total cerebral white matter abnormality score, infants were divided into three groups. These groups included the following: no abnormalities (composite score 5–7, $n = 23$), mild abnormalities (composite score 8–10, $n = 54$), and moderate to severe abnormalities (composite score > 10, $n = 17$). Further details about the scoring of the MR data can be found in an online supplement to Woodward et al. (2006).

Measures of family social background and parenting. Five measures of children's social and family background were included in the analysis. The first was maternal age at birth. The second was SES, as measured by the Elly–Irving Socioeconomic Index (Elly & Irving, 2003). Families were assigned an SES level ranging from 1 (*professional*) to 6 (*unemployed*) based on the profession of the highest earning parent. This scale has been standardized for use with New Zealand families and has good predictive validity (Fergussen & Woodward, 2000). The third measure was maternal anxiety or depression as assessed by the Hospital Anxiety and Depression Scale (HADS; Zigmond & Snaith, 1983). This 14-item measure consists of statements such as "I feel as if I'm slowed down" and "I get sudden feelings of panic," which mothers rated on a scale of 0 (*never*) to 3 (*most of the time*). Scores were reverse coded where necessary and summed to produce a composite index of depression and anxiety. The HADS has been shown to have adequate reliability and validity (Herrmann, 1997; Mykletun, Stordal, & Dahl, 2001). Internal consistency, assessed using Cronbach's alpha, was .80 in this sample.

In addition to the child behaviors, two parenting behaviors, rated on the basis of the videotaped parent–child interaction obtained at age 2 years, were selected for inclusion in this analysis. These scales included parental sensitivity and intrusiveness and were rated on a 5-point Likert scale, where 1 represented a low degree of sensitivity/intrusiveness and 5 represented a high degree of these behaviors. *Parental sensitivity* provided a measure of the verbal and nonverbal support that a parent showed toward their child. Parents who showed supportive behaviors such as leaning in toward their child when they were having more difficulty, being involved and encouraging, providing assistance in a timely fashion, pacing the interaction so that the child remained engaged, or allowing the child to explore the puzzles while anticipating when they required more support were given higher ratings on this scale. *Parental intrusiveness* measured the extent of excessively or inappropriately controlling actions or commands. Parents who were directive in their instructions, who were interfering, or who tended to take over the completion of a task without allowing the child time to explore or tackle the tasks independently were rated higher on this measure. These measures were based on previously established schemes (Belsky et al., 2005). Interrater reliabilities, based on Pearson product moment correlations, were .84 for parental sensitivity and .68 for intrusiveness.

Finally, a measure of children's exposure to early intervention was also included. This measure was based on parent-reported intervention experience up to age 4 years. Only those children who attended early intervention beyond 6 months were considered to have received intervention, as a large proportion of children underwent monitoring for potential intervention prior to this age. Intervention was generally multidimensional and targeted at both motor and cognitive functioning.

Statistical Methods

Between-group differences in the medical and social characteristics of the very preterm, extremely preterm, and full-term groups were examined using analysis of variance (ANOVA) for continuous variables and the chi-square test of significance for categorical variables. Following this, data analysis was completed in three parts, corresponding to each study aim. First, ANOVA was used to examine mean differences on individual measures of regulatory competence across the three groups of children (extremely preterm, very preterm, and full term). In general, Kolmogorov–Smirnov tests on the data from these measures showed that distributions were skewed. Thus, these analyses were rerun using nonparametric Kruskal–Wallace tests. The results from these tests were consistent with those reported for the ANOVAs. This analysis was then extended to control for the effects of between-group differences in SES using analysis of covariance.

Second, in order to examine the extent to which a composite measure of children's regulatory competence across contexts might be created, bivariate Pearson product moment correlations were used to examine the concordance between multiple measures of regulation across the whole sample, the group of children born very/extremely preterm, and the group of children born full term at each age studied. Following this, principal components analysis (PCA) was used to create a regulation composite. Between-group differences in performance on this composite across the two time points studied were then examined using a repeated measures ANOVA.

Third, in order to examine the relationship between a number of social and medical factors and later regulation, a general estimating equation (GEE) regression model was fitted to the regulation composites across the two ages for the very preterm and extremely preterm groups only. This analysis included consideration of fluctuations in children's regulatory competence over time, while taking into account the correlations between repeated observations (Hanley, Negassa, Edwardes, & Forrester, 2003). A p value of .05 was used for significance testing, apart from the analysis for the final regression model, where a significance level of $p < .10$ was used to determine whether each predictor should be included in or excluded from the model.

Results

Characteristics of the Sample

Table 1 provides a descriptive profile of the clinical, social background and family characteristics of children in each study group. Inspection of Table 1 shows that decreasing gestational age at birth was associated with increased risks of medical complications, including lower birth weight ($p < .001$), IUGR ($p < .05$), moderate-to-severe cerebral white matter abnormalities ($p < .001$), chronic lung disease ($p < .001$), sepsis ($p < .001$), PDA ($p < .001$), and NEC ($p < .001$). Multiple births were also more common in the very preterm and extremely preterm groups ($p < .001$) as was exposure to early intervention ($p < .001$).

Consideration of the family social background characteristics of study children also revealed that children born extremely preterm or very preterm were characterized by higher rates of socioeconomic disadvantage ($p < .001$), maternal educational underachievement ($p < .01$), and maternal smoking during pregnancy ($p < .05$). There was also a tendency for mothers of these infants to have higher depression–anxiety scores than mothers of term-born control children at 2 years, although this did not reach statistical significance. Observation of parenting at age 2 years showed that parents of children born extremely preterm and very preterm appeared less adept at anticipating cues, responding to signals in a timely fashion ($p < .05$), and that they behaved more intrusively ($p < .01$) during problem solving.

Emotional and Behavioral Regulation at Ages 2 and 4 Years

Table 2 describes the early emotional and behavioral regulatory abilities of extremely preterm, very

Table 1
Infant and Family Characteristics of Children Born Extremely Preterm, Very Preterm, and Full Term

	Gestational age group				
Measure	Extremely preterm (n = 39)	Very preterm (n = 55)	Full term (n = 103)	F/χ^2	p
Infant clinical characteristics (birth)					
Mean (*SD*) gestational age	25.51 (1.45)	29.45 (1.32)	39.59 (1.10)	2292.04	< .001
Mean (*SD*) birth weight	809.05 (236.86)	1,254.82 (225.80)	3,580.21 (399.93)	1424.99	< .001
% Male	48.7	50.9	52.5	0.17	.92
% Twin	30.8	36.4	1.9	34.6	< .001
% Intrauterine growth restriction	12.8	3.6	1.0	9.87	.01
% Moderate to severe white matter abnormality	28.9	10.9	—	28.92	< .001
% Chronic lung disease	64.1	12.7	—	83.95	< .001
% Suspected or proven sepsis	80.6	47.3	—	10.08	< .001
% Suspected or proven necrotizing enterocolitis	15.4	1.8	—	6.09	.01
% Patent ductus arteriosus	66.7	30.9	—	11.76	< .001
% Maternal antenatal steroids	89.7	78.2	—	2.16	.14
Family background/socioenvironmental characteristics					
% Early intervention beyond 6 months	89.5	61.8	1.0	115.77	< .001
Mean (*SD*) maternal age (birth)	31.05 (6.15)	30.64 (4.76)	31.02 (4.59)	0.12	.89
% Semi/unskilled socioeconomic status (birth)	38.5	36.4	11.2	17.87	< .001
% Mother left school <16 years (birth)	38.5	41.8	19.8	10.57	.005
% New Zealand European (birth)	92.3	96.4	92.9	0.90	.64
% Maternal smoking during pregnancy	33.3	36.4	15.5	9.90	.01
Mean (*SD*) maternal depression–anxiety score (2 years)	9.73 (5.15)	9.39 (5.26)	8.30 (5.20)	1.33	.26
Mean (*SD*) observed parent sensitivity (2 years)	3.18 (0.87)	3.48 (0.77)	3.55 (0.78)	3.08	.05
Mean (*SD*) observed parent intrusiveness (2 years)	1.69 (0.84)	1.45 (0.57)	1.31 (0.45)	5.96	.003

Table 2

Early Emotional and Behavioral Regulation of Children Born Extremely Preterm, Very Preterm, and Full Term at Ages 2 and 4 Years

	Gestational age group					
	Extremely preterm ($n = 39$)	Very preterm ($n = 56$)	Full-term control ($n = 103$)	F	p	Effect size
2 years						
Mean (*SD*) observed positive affect	2.01 (0.85)	1.95 (0.82)	2.25 (0.84)	3.56	.08	.29; .36
Mean (*SD*) observed negative affect	1.55 (0.72)	1.38 (0.51)	1.46 (0.58)	0.90	.41	.15; .14
Mean (*SD*) observed persistence	3.20[a] (0.94)	3.46 (0.73)	3.63 (0.71)	4.50	.01	.56; .24
Mean (*SD*) observed transition	3.34 (0.84)	3.23 (0.93)	3.51 (0.80)	2.02	.14	.21; .33
Mean (*SD*) tester-rated regulation	3.20[a] (0.84)	3.55 (0.76)	3.72 (0.78)	6.17	.003	.66; .22
Mean (*SD*) parent-reported regulation	4.09 (0.47)	4.06 (0.62)	4.17 (0.34)	1.60	.21	.22; .15
4 years						
Mean (*SD*) observed positive affect	2.54 (0.84)	2.41 (0.78)	2.48 (0.85)	0.34	.71	.07; .09
Mean (*SD*) observed negative affect	1.37 (0.65)	1.23 (0.33)	1.22 (0.45)	1.63	.20	.30; .01
Mean (*SD*) observed persistence	3.84[b] (0.94)	4.31 (0.71)	4.26 (0.73)	5.04	.007	.53; .07
Mean (*SD*) observed transition	4.64 (0.48)	4.75 (0.50)	4.70 (0.49)	0.61	.54	.12; .10
Mean (*SD*) tester-rated regulation	3.45[a] (0.72)	3.71 (0.74)	3.80 (0.73)	3.33	.04	.48: .12
Mean (*SD*) parent-reported regulation	4.07[a] (0.52)	4.28 (0.58)	4.39 (0.39)	6.42	.002	.75; .25

Note. Effect size (d) comparing (a) extremely preterm and full-term groups and (b) very preterm and full-term groups based on Cohen's criteria: .2 = small, .5 = medium, and .8 = large.

[a] Mean score for extremely preterm group is significantly different from that for full-term group ($p < .05$). [b] Mean score for extremely preterm group is significantly different from that for very preterm and full-term groups ($p < .05$).

preterm, and full-term infants across a range of measures administered at age 2 years and then repeated at age 4 years. These measures included independent ratings of children's observed emotional behavior (positive and negative) and regulatory control during a 10- to 15-min structured parent–child interaction at the ages of both 2 and 4 years, tester ratings of children's emotional regulation during cognitive assessments, and parental ratings of children's emotional regulation on the Emotion Regulation Checklist. Associations between gestational age at birth and children's subsequent emotional and behavioral regulatory skills were examined using one-way ANOVA, with Tukey tests for honestly significant differences used for post hoc, between-group contrasts (e.g., extremely vs. very preterm). The results of this analysis are presented in Table 2 and show the following:

1. At age 2 years, significant associations were found between children's gestational age at birth and their subsequent persistence and sustained attention during a series of problem-solving tasks with their mother ($p < .05$), as well as their observed emotional regulation during testing ($p < .01$). Specifically, decreasing gestational age was associated with poorer emotional regulation, attention, and behavioral control across two independent observational assessments at age 2 years. There was also a tendency for children born at lower gestational ages to be characterized by lower levels of positive affect than term-born infants ($p < .10$). Post hoc Tukey tests showed that children in the extremely preterm group were significantly less likely than those in the full-term group to show persistence during the problem-solving interactions ($p < .01$). They were also rated by examiners as showing less regulatory competence during testing, relative to the full term ($p < .01$), group and there was a tendency for them to show less positive affect while engaging in problem solving ($p < .10$). Effect sizes for significant differences between the extremely preterm and full-term groups at this age were small to moderate (.36–.66), whereas those for the very preterm relative to the full-term group were small or negligible (.15–.36; Cohen, 1992). No significant associations were found between gestational age and children's observed negative affect and ability to transition between different tasks during play with their primary caregiver or in terms of parental perceptions of emotional regulation.
2. At age 4 years, a significant relationship was also found between gestational age and persistence

> and engagement ($p < .05$) during parent–child interaction. A significant relationship was also shown between gestational age group and tester-rated emotional regulation ($p < .05$). Additionally, there was a significant relationship between gestational age and parent-reported emotional regulation ($p < .01$). Post hoc Tukey tests showed that children in the extremely preterm group showed less persistence than both the very preterm and full-term groups across the problem-solving tasks ($p < .05$) and were rated by examiners as being less regulated during testing ($p < .05$). They were also rated by parents as having less ability to regulate their emotional responses and behavior than the full-term ($p < .01$) and, to a lesser extent, the very preterm ($p < .10$) group. Effect sizes for the differences between the extremely preterm and full-term groups were small to moderate (.30–.73), whereas those for the differences between the very preterm and full-term groups were generally negligible (.01–.24). No significant associations were found between gestational age and the extent of children's observed positive or negative affect or their ability to transition smoothly between tasks when interacting with a primary caregiver.

Given the significant differences in the socioeconomic backgrounds of the full-term and preterm groups, these analyses were rerun controlling for household SES. Findings from this analysis were consistent with those reported above, with all significant associations remaining even after taking family SES into account.

Collectively, these findings revealed a generally significant pattern of differences between children of varying gestational ages in their ability to regulate themselves during testing and in their persistence during interactions with their primary caregivers at ages 2 and 4 years. In addition, findings suggest that children born extremely preterm tend to be characterized by more pervasive dysregulation than their very preterm and full-term peers, with effect sizes for significant differences between extremely preterm and full-term children tending to be larger than those between the very preterm and full-term groups.

Development of Self-Regulation Between the Ages of 2 and 4 Years

The general pattern of associations between gestational age at birth and children's self-regulatory abilities shown in Table 2 raises the issue of how these skills may develop and change over time in children exposed to varying degrees of biological risk. This question is important because it may help clarify whether observed regulatory impairments reflect a persistent deficit in regulation, a transient delay, or a divergent developmental course for all or some children born preterm. This question was addressed by the following analyses.

First, we examined the extent to which the observational and parent report measures of children's emotional and behavioral regulation might reflect a common regulation factor at each assessment point. Such a composite measure would, theoretically, offer a more robust measure, spanning a range of salient aspects of children's emerging regulation observed across multiple contexts. Methodologically, such a composite would also enable a clearer examination of the development of regulation across time, as well as the factors influencing the development of children's regulatory competence over early childhood. This analysis was conducted in three steps. First, intercorrelations between measures at each time point and across time were examined. Most of these correlations were significant ($p < .05$), with evidence of both concordance across raters and situations, as well as some indication of stronger correlations between measures becoming evident with age, particularly with respect to parent reports of emotional regulation and independent ratings of child behavior during parent–child interaction.

Given the presence of significant associations between individual measures, scores of the full sample were subjected to PCA using an eigenvalue criterion of 1. Results from this analysis supported the presence of a single, common regulation factor at the ages of both 2 and 4 years. At age 2 years, the variance explained by this factor was 39.81%. Although parent-reported emotion regulation showed a much lower loading at age 2 years (.21) than did other factors (.58–.80), it was a stronger indicator of the factor at age 4 years (.53). At age 4 years, this factor explained 44.5% of the variance in outcome, with factor loadings ranging from .40 to .85. To assess whether this pattern differed for preterm and full-term children, the PCA was rerun separately for (a) children in the full-term group and (b) children in the preterm group. For the full-term group, a single factor solution was obtained, explaining 40.2% of variance at age 2 years and 47% at age 4 years. In the very preterm group, the parent-reported ratings on the Emotion Regulation Checklist comprised a separate factor at age 2 years, although all other measures had loadings of .55–.77. At age 4 years, all measures loaded consistently on one factor, which explained 43.35% of the variance in outcome. In general, these analyses provided statistical support for the formation of a composite measure of child regulation. Although for extremely and very preterm children, the parent report measure at age 2 years was not correlated with other measures, by age 4 years it was consistently associated with other measures. In addition, because it draws on the perspective of the parent, the inclusion of this measure provides for the description of regulatory competence across different contexts. Therefore, the decision was made to include it in our summary scores of regulatory competence.

To create the composite measures, children's individual scores on each of the measures collected across the two assessment time points were summed to form a composite measure of children's overall self-regulation at ages 2 and 4 years. Scores for child negative affect, as rated from the problem-solving activities with parents, were reverse coded so that all scores reflected a tendency toward better regulation of behavior and affect. Individual item scores were not weighted because all measures were scored on a common metric from 1 to 5. Cronbach's alphas for the scales were .68 at age 2 years and .71 at age 4 years.

Figure 2 illustrates the relationship between children's regulation at age 2 and 4 years for each of the three study groups. Examination of children's composite regulation scores over time showed that, across all groups, children's scores improved between the ages of 2 and 4 years. However, at both time points, children born extremely preterm were characterized by consistently lower scores than their full-term peers. In contrast, children in the very preterm group showed lower scores than full-term children at age 2 years but by age 4 years, achieved similar levels of self-regulation to children born full term. This pattern of findings was confirmed by a repeated-measures ANOVA with a 3 (gestational age group) × 2 (age assessed) design. Results demonstrated a significant age effect, $F(1, 182) = 236.25$, $p < .05$, showing that self-regulation improved across early childhood in all three groups. A significant group effect was also found, $F(2, 182) = 6.10$, $p < .01$, confirming earlier between-group differences in the ability of children to regulate themselves. Post hoc analyses confirmed that the extremely preterm group differed significantly from the full-term group in terms of their self-regulation across both ages ($p < .01$), although differences between the very preterm group and full-term group were at a trend level at age 2 ($p < .10$) and not significant at age 4 years ($p > .10$). In fact, at this age, post hoc testing showed that there was also a statistical trend toward lower performance in the extremely preterm group relative to the very preterm group ($p < .10$).

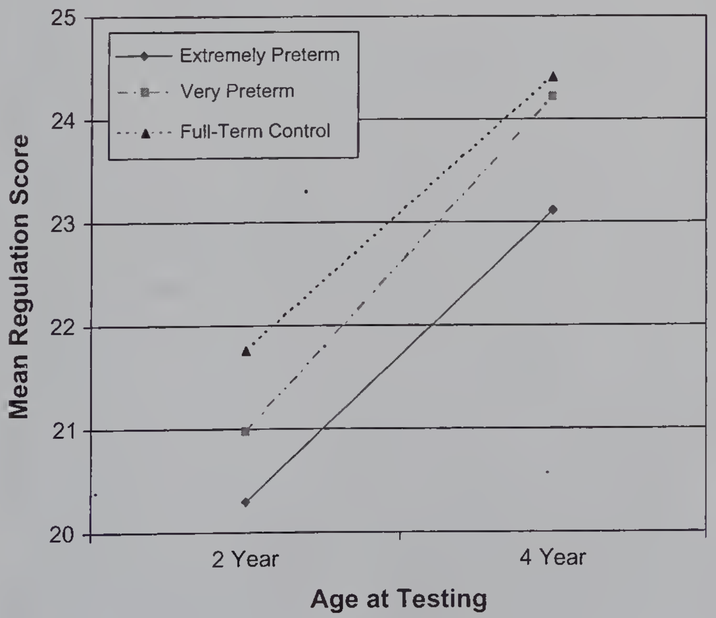


Figure 2. Development of regulation between 2 and 4 years in children born extremely preterm, very preterm, and full term.

Clinical and Sociofamilial Factors Influencing the Development of Self-Regulation in Children Born Preterm

The above findings suggest that decreasing gestational age at birth places children at increasing risk for regulatory difficulties, with extremely preterm children appearing to be especially vulnerable to persistent impairments relative to older preterm-born children and full-term-born children. However, it is also possible that these between-group differences in the development of self-regulatory skills among extremely and very preterm children may be partially or wholly explained by other factors correlated with gestational age at birth. For example, there is clear evidence from this (Table 1) and previous studies that children born extremely and very preterm often differ on a range of other clinical and sociofamilial factors that may also influence their longer term developmental outcome. Specifically, comparisons between the two preterm groups showed that children born extremely preterm had higher rates of perinatal medical complications, including lower birth weight ($p < .001$), chronic lung disease ($p < .001$), IUGR ($p < .10$), PDA ($p < .001$), sepsis ($p < .001$), NEC ($p < .05$), and cerebral white matter abnormalities on term MR imaging ($p < .001$). In addition, children born extremely preterm received a greater number of hours of intervention ($p < .05$). Although families of children born extremely preterm and very preterm did not differ in terms of SES, ethnicity, maternal age, smoking during pregnancy, or maternal depression, parents of children born extremely preterm appeared less sensitive to their children's cues ($p < .05$) and more intrusive ($p < .05$) during problem solving at age 2 years.

To examine the influence of these medical and social risk factors on preterm children's ability to regulate their emotions and behavior, a GEE regression model was fitted with regulatory competence at 2 and 4 years as the dependent variables. This method enabled the prediction of children's regulation across both time points within the same regression analysis. Each clinical and family background factor outlined in Table 1 was entered into the model in a forward and backward manner to test for significant relationships with later regulation. In addition, age was entered into the model to test for any differences across the two ages studied. The analysis was confined to the group of children born preterm. The final main effects model is shown in Table 3. As can be seen from this model, two perinatal factors, the presence of moderate to severe white matter abnormalities on term MR imaging ($p < .05$) and maternal antenatal steroid exposure ($p < .001$), were significantly related to later regulatory competence. Specifically, the administration of antenatal steroids played a protective role; however, moderate to severe white matter abnormalities on MR were associated with increased risk of poor regulation ($p < .05$). In terms of sociofamilial factors, parental sensitivity at age 2 years was an important independent predictor of children's later regulatory competence ($p < .05$). Finally, age was also found to be a significant predictor of children's regulation ($p < .001$), supporting earlier findings that children's regulatory abilities improved from 2 to 4 years of age. No significant interactions were found between each of the covariates and age, suggesting that these influences were not age dependent but were predictive of self-regulation across both ages.

Table 3
Predictors of Regulation in Children Born Extremely Preterm and Very Preterm at Ages 2 and 4 Years

	β	*SE*	χ^2	*p*
Extremely preterm birth	0.48	0.47	1.03	.31
Moderate to severe white matter abnormalities (term)	−1.62	0.63	6.54	.01
Antenatal steroids (birth)	1.44	0.47	9.43	.002
Maternal supportive presence during problem solving (2 years)	0.98	0.26	14.20	< .001
Time (2 or 4 years)	1.65	0.14	143.06	< .001

Discussion

The present study extends on previous research concerned with the impact of preterm birth on children's social and emotional development. Existing studies have primarily focused on the outcomes of school-aged children and adolescents born preterm, with little consideration given to the early development of emotional and behavioral competence in this high-risk population. Furthermore, little consideration has been given to the role of gestational age at birth, particularly extreme prematurity, or the explanatory role of other biological and sociocontextual factors. This prospective longitudinal study addressed these important gaps by examining the early emotional and behavioral regulation of children born extremely preterm, very preterm, and full term during early childhood. Strengths of the study included the relatively large sample size, the use of a multimethod measurement approach to the assessment of regulation, the use of repeated measures over time, inclusion of neuroanatomical MR measures, and detailed observational and parent report measures of children's social and familial experiences during infancy and early childhood. The major study findings are discussed below.

Extending on previous research demonstrating emotional and behavioral disturbances in older children born very preterm (Botting et al., 1997; Breslau et al., 1996; Klebanov et al., 1994; McCormick et al., 1996; Sykes et al., 1997), this study revealed the presence of pervasive differences between toddlers and preschool children born extremely preterm, very preterm, and full term on a range of measures of early emotional and behavioral development. At age 2 years, children born extremely preterm showed poorer emotional and behavioral regulation during cognitive testing and were less persistent in completing a series of problem-solving tasks. A generally similar pattern of emotional and behavioral differences emerged at age 4 years. However, by age 4 years, parents of children born extremely preterm were also more likely to report difficulties in emotional regulation, perhaps indicating that these difficulties were becoming more apparent to parents as children matured and parental expectations increased. Ratings of positive and negative affect during problem solving with parents did not differ significantly between groups. This may reflect the nature of the task, which was designed to be enjoyable and thus not particularly stressful for children or parents, perhaps resulting in fewer negative behaviors than a more structured and demanding task such as a cleanup task.

There is some overlap between these findings and those of studies examining the neurobehavioral outcomes of young children born very preterm. Based on experimental and observational measures, several studies have indicated that attention is an area of difficulty for children born very preterm (Caravale, Tozzi, Albino, & Vicari, 2005; Rose, Feldman, Wallace, & McCarton, 1989; Vicari, Caravale, Carlesimo, Casadei, & Allemand, 2004). For example, Landry and her colleagues (Landry & Chapieski, 1988; Landry et al., 1996) found that children born very preterm with a high degree of perinatal risk were less able to shift their attention effectively from one toy to another when examining toys alone or during interaction with their parents at age 6 months. In the present study, there were no differences in children's ability to transition from one task to another. Instead, children differed in the levels of persistence or sustained attention they showed during problem solving. Studies that have examined executive abilities in older children born preterm have also produced conflicting results regarding the specific areas of attention that may be disrupted, with some suggesting that sustained attention and inhibitory control may be areas of difficulty (Bohm, Smedler, & Forssberg, 2004; Katz et al., 1996; Korkman, Liikanen, & Fellman, 1996), whereas others indicate that attentional flexibility or shifting may be impaired (Bayless & Stevenson, 2007).

There is some possibility that different attentional impairments may become more apparent at different developmental stages, thereby explaining some of the variation in findings. However, these differences in findings may also be related to the paradigms employed. In this study, we only recorded quality of transition between tasks. We did not examine the number of times that children's attention shifted to different toys within each of the tasks. Also, although Landry and Chapieski (1988) found that children at high medical risk made fewer shifts between toys, they also found that children spent less time engaged with toys overall, suggesting that these children may not have shifted attention simply because their attention was focused elsewhere. In this case, it is difficult to clearly separate out the different facets of attention. With executive function becoming a popular area of interest for developmental psychologists and more precise tools for the measurement of attention emerging, these issues should be resolved. Regardless, this study is consistent with earlier studies of children born preterm in suggesting that difficulties in the regulation of attention and emotion are apparent at an early age (Garcia Coll et al., 1988; Lowe, Woodward, & Papile, 2005; Ohgi, Takahashi, Nugent, Arisawa, & Akiyama, 2003; Segal et al., 1995).

Examination of the development of children's behavioral and emotional regulation between the ages of 2 and 4 years showed that, across all groups, children showed developmental gains in overall regulation. Among the two high-risk groups (very preterm and extremely preterm), improvements were most marked in the older gestation, very preterm group. In contrast, children born extremely preterm, although showing developmental gains, continued to score significantly less well than their full-term peers. Continued follow-up of these children to school age will be important to help clarify the longer term implications of these findings.

These between-group differences in the development of self-regulation raise the important question of why some children born preterm, especially those born at lower gestational ages, tend to fare less well than others. An examination of the extent to which between-group differences in the development of regulation among children born extremely and very preterm could be explained by a range of medical and social factors correlated with gestational age revealed that both biological and sociofamilial factors made important contributions. Medical factors that emerged as important for the development of later regulation in this study were maternal antenatal steroid exposure, which was associated with better regulation, and moderate to severe white matter abnormality, which was associated with a poorer outcome. The finding of increased regulatory competence in children whose mothers were administered antenatal steroid therapies is intriguing and warrants further replication and investigation.

Once the presence of white matter abnormality was statistically controlled for, the effect of gestational age was no longer significant. This suggests that it may not be extreme prematurity per se that places these infants at risk for observable regulatory difficulties. Rather, findings suggest that cerebral white matter abnormalities, including white matter volume loss, cystic abnormalities, and immature patterns of myelination, may place constraints on the development of regulation. These findings are largely consistent with those of McGrath et al. (2005), who found that children born preterm with no medical or neurological complications showed similar levels of attentional regulation to their full-term peers. In contrast, those who were characterized by higher levels of neurological and medical risk showed greater impairments in attentional regulation. An advantage of the present study is that we examined independent risk factors in determining outcome, as opposed to relying on a composite medical risk score. This allowed us to isolate early neurological abnormality as being individually important for later regulatory development. Indeed, the present study contributes to a growing body of literature showing that early cerebral abnormalities, as determined by MR, may be a sensitive predictor of later impairment in children born very preterm (Peterson et al., 2003; Stewart et al., 1999; Woodward et al., 2006).

As others have noted (Nadeau et al., 2003; Taylor, Minich, Bangert, Filipek, & Hack, 2004), it does seem that the neurological effects of being born very early in these smallest infants have ongoing implications for longer term brain and behavioral development. The principal region of injury in the premature infant is the periventricular zone, which is an important area of neuronal integration and cortical projection. Damage in this region has been associated with disrupted myelination and organization of important fiber tracts connecting subcortical and cortical regions (Counsell et al., 2003; Huppi et al., 2001). As the brain becomes more integrated and there is increased modulation of lower level thalamic and cingulate processes by higher level frontal areas, children gradually develop a better ability to inhibit inappropriate responses, flexibly orient their attention, and organize their emotional responses (Diamond, 2002; Posner & Rothbart, 2000; Thompson, 1994). It may be that the disruption of neural development, caused by early insult or alterations to white matter tracts, interrupts the integrated coordination of these processes, thereby hindering children's ability to regulate effectively.

In addition to biological factors, the extent to which parents were responsive and sensitive to the needs of their child was also an important predictor of regulatory competence. Findings showed that parents of extremely preterm children had more difficulty scaffolding the interactions around cues from their children and providing well-timed support. Rather, they tended to become more intrusive during problem solving at age 2 years. These findings are consistent with mainstream developmental studies highlighting the importance of parental sensitivity for child cognitive and social development, as well as studies of children born at high biological risk (Landry, Smith, & Swank, 2003, 2006). Because ratings of parental and child behavior were taken during an interactive task, the direction of these effects is not clear. It is likely, however, that parent and child characteristics are interacting in a bidirectional manner. Children who have difficulties regulating their arousal and affect will require a higher degree of facilitation, soothing, and calming by parents. Coupled with this, signals from these children may be more ambiguous. For instance, children born very preterm have been found to exhibit more mixed cues with their caregivers (Eckerman, Hsu, Molitor, Leung, & Goldstein, 1999). Parents may react by becoming more intrusive in order to evoke a response from the child. This may be overstimulating for a child who is already trying to compensate for disrupted homeostasis, thereby setting up aversive social interactions for both parties. This will have ongoing effects on the development of the mutually rewarding relationship between parent and child.

Difficulties in the regulation of affect and attention have a number of implications for ongoing development. Subtle difficulties in regulation and emotional control may well become more apparent with the complex and challenging demands of the classroom, impacting on learning, self-esteem, and social development. Thus, early regulatory support may be an important target area for early intervention. Some support for this exists in the Newborn Individualized Developmental Care and Assessment Program, an intervention aimed at regulating arousal and improving signal recognition (i.e., sensitivity) of parents of compromised preterm babies. This intervention showed discriminate effects on later child behavior and quality of mother–child interaction (Kleberg, Westrup, & Stjernqvist, 2000). Furthermore, new exciting findings show that these changes may be accompanied by improved neurobehavioral functioning but also more mature frontal brain structure development (Als et al., 2004).

Although this study had a number of methodological strengths, including the use of multiple measures across time, a major challenge with this, as with other similar studies of very young children, is the lack of well-validated measures of early self-regulation. Relatedly, the need for longitudinal measures repeated across time possibly obscures important changes in development. It is possible that aspects of self-regulation applicable at age 2 years may become less salient with age. It may be that 2-year-olds are more likely, in general, to show extreme levels of affect. By age 4 years, such difficulties may manifest more in the regulation of motivation and attention, constructs that clearly differentiated children at this age. Further assessment of more specific and subtle behavioral differences that reflect poor regulation at various ages as well as the development of measures that can capture these differences remains an ongoing challenge for developmental research.

Nonetheless, this study does provide initial evidence of early and stable deficits in self-regulation during early childhood among children born preterm. It also highlights the importance of both neurological and social influences in accounting additively for differences in the self-regulatory abilities of children born extremely preterm, very preterm, and full term. Neonatal strategies aimed at minimizing the neurological impacts of being born too early and early intervention strategies that support parent's ability to support the developmental needs of their infants in a sensitive manner are likely to be important in reducing ongoing behavioral risks. Although it represents a significant step toward understanding the specific mechanisms and processes that lead to poor outcomes after preterm birth, future research through longer periods of follow-up is needed to ascertain the relevance of self-regulation, in relation to other developing skills and capacities, for long-term cognitive and socioemotional development in this vulnerable population.

References

Als, H., Duffy, F. H., McAnulty, G. B., Rivkin, M. J., Vajapeyam, S., Mulkern, R. V., et al. (2004). Early experience alters brain function and structure. *Pediatrics, 113*, 846 – 857.

Anderson, P., Doyle, L. W., & the Victorian Infant Collaborative Study Group. (2003). Neurobehavioral outcomes of school-age children born extremely low birth weight or very preterm in the 1990s. *Journal of the American Medical Association, 289*, 3264 – 3272.

Bayless, S., & Stevenson, J. (2007). Executive functions in school-age children born very prematurely. *Early Human Development, 83*, 247 – 254.

Bayley, N. (1993). *The Bayley Scales of Infant Development – Revised*. New York: Psychological Corporation.

Belsky, J., Jaffee, S. R., Sligo, J., Woodward, L., & Silva, P. A. (2005). Intergenerational transmission of warm-sensitive-stimulating parenting: A prospective study of mothers and fathers of 3-year-olds. *Child Development, 76*, 384 – 396.

Bhutta, A. T., Cleves, M. A., Casey, P. H., Cradock, M. M., & Anand, K. J. (2002). Cognitive and behavioral outcomes of school-aged children who were born preterm: A meta-analysis. *Journal of the American Medical Association, 288*, 728 – 737.

Bohm, B., Smedler, A. C., & Forssberg, H. (2004). Impulse control, working memory and other executive functions in preterm children when starting school. *Acta Paediatrica, 93*, 1363 – 1371.

Botting, N., Powls, A., Cooke, R. W., & Marlow, N. (1997). Attention deficit hyperactivity disorders and other psychiatric outcomes in very low birthweight children at 12 years. *Journal of Child Psychology & Psychiatry & Allied Disciplines, 38*, 931 – 941.

Breslau, N., Brown, G. G., DelDotto, J. E., Kumar, S., Ezhuthachan, S., Andreski, P., et al. (1996). Psychiatric sequelae of low birth weight at 6 years of age. *Journal of Abnormal Child Psychology, 24*, 385 – 400.

Brown, N. C., Doyle, L. W., Bear, M. J., & Inder, T. E. (2006). Alterations in neurobehavior at term reflect differing perinatal exposures in very preterm infants. *Pediatrics, 118*, 2461 – 2471.

Calkins, S. D. (1994). Origins and outcomes of individual differences in emotion regulation. *Monographs of the Society for Research in Child Development, 59*, 53 – 72.

Caravale, B., Tozzi, C., Albino, G., & Vicari, S. (2005). Cognitive development in low risk preterm infants at 3-4 years of life. *Archives of Disease in Childhood Fetal & Neonatal Edition, 90*, F474 – F479.

Chase-Lansdale, P. L., Brooks-Gunn, J., & Zamsky, E. S. (1994). Young African-American multigenerational families in poverty: Quality of mothering and grandmothering. *Child Development, 65*, 373 – 393.

Cohen, J. (1992). A power primer (quantitative methods in psychology). *Psychological Bulletin, 112*, 155 – 159.

Cole, P. M., Michel, M. K., & Teti, L. O. (1994). The development of emotion regulation and dysregulation: A clinical perspective. *Monographs of the Society for Research in Child Development, 59*, 73 – 100.

Colman, R., Hardy, S., Albert, M., Raffaelli, M., & Crocket, L. (2006). Early predictors of self-regulation in middle childhood. *Infant and Child Development, 15*, 421 – 437.

Cooke, R. W. (2004). Health, lifestyle, and quality of life for young adults born very preterm. *Archives of Diseases in Childhood, 89*, 201 – 206.

Counsell, S. J., Allsop, J. M., Harrison, M. C., Larkman, D. J., Kennea, N. L., Kapellou, O., et al. (2003). Diffusion-weighted imaging of the brain in preterm infants with focal and diffuse white matter abnormality. *Pediatrics, 112*, 1 – 6.

Dahl, R. E. (2003). The development of affect regulation: Bringing together basic and clinical perspectives. *Annals of the New York Academy of Sciences, 1008*, 183 – 188.

DeGangi, G. (2000). *Pediatric disorders of regulation in affect and behavior.* San Diego, CA: Academic Press.

Delobel-Ayoub, M., Kaminski, M., Marret, S., Burguet, A., Marchand, L., N'Guyen, S., et al. (2006). Behavioral outcome at 3 years of age in very preterm infants: The EPIPAGE study. *Pediatrics, 117*, 1996 – 2005.

Diamond, A. (2002). Normal development of the PFC from birth to young adulthood: Cognitive functions, anatomy and biochemistry. In D. T. Stuss & R. T. Knight (Eds.), *Principles of frontal lobe function.* (pp. 466 – 503). Oxford, UK: Oxford University Press.

Doyle, L. W., Casalaz, D., & the Victorian Infant Collaborative Study Group. (2001). Outcome at 14 years of extremely low birthweight infants: A regional study. *Archives of Disease in Childhood Fetal & Neonatal Edition, 85*, F159 – F164.

Duffy, F. H., Als, H., & McAnulty, G. B. (1990). Behavioral and electrophysiological evidence for gestational age effects in healthy preterm and full term infants studied two weeks after expected due date. *Child Development, 61*, 271 – 286.

Eckerman, C. O., Hsu, H. C., Molitor, A., Leung, E. H., & Goldstein, R. F. (1999). Infant arousal in an en-face exchange with a new partner: Effects of prematurity and perinatal biological risk. *Developmental Psychology, 35,* 282–293.

Eckerman, C. O., Oehler, J. M., Medvin, M. B., & Hannan, T. E. (1994). Premature newborns as social partners before term age. *Infant Behavior and Development, 17,* 55–70.

Edgin, J. O., Inder, T. E., Anderson, P. J., Hood, K. M., Clark, C. A. C., & Woodward, L. J. (2008). Executive function in preschool children born very preterm: Relationship with early white matter pathology. *Journal of the International Neurological Society, 14,* 90–101.

Eisenberg, N., Guthrie, I. K., Fabes, R. A., Shepard, S., Losoya, S., Murphy, B. C., et al. (2000). Prediction of elementary school children's externalizing problem behaviors from attentional and behavioral regulation and negative emotionality. *Child Development, 71,* 1367–1382.

Elly, W. B., & Irving, J. C. (2003). The Elly-Irving socioeconomic index: 2001 census revision. *New Zealand Journal of Educational Studies, 38,* 3–17.

Farooqi, A., Hagglof, B., Sedin, G., Gothefors, L., & Serenius, F. (2007). Mental health and social competencies of 10- to 12-year-old children born at 23 to 25 weeks of gestation in the 1990s: A Swedish national prospective follow-up study. *Pediatrics, 120,* 118–133.

Fergussen, D., & Woodward, L. J. (2000). Family socioeconomic status at birth and rates of university participation. *New Zealand Journal of Educational Studies, 35,* 25–36.

Forcada-Guex, M., Pierrehumbert, B., Borghini, A., Moessinger, A., & Muller-Nix, C. (2006). Early dyadic patterns of mother-infant interactions and outcomes of prematurity at 18 months. *Pediatrics, 118,* e107–e114.

Foster-Cohen, S., Champion, P., Edgin, J. O., & Woodward, L. J. (2007). Early language development of children born very premature. *Journal of Child Language, 34,* 655–675.

Garcia Coll, C. T., Emmons, L., Vohr, B. R., Ward, A. M., Brann, B. S., Shaul, P. W., et al. (1988). Behavioral responsiveness in preterm infants with intraventricular hemorrhage. *Pediatrics, 81,* 412–418.

Gardner, F., Johnson, A., Yudkin, P., Bowler, U., Hockley, C., Mutch, L., et al. (2004). Behavioral and emotional adjustment of teenagers in mainstream school who were born before 29 weeks' gestation. *Pediatrics, 114,* 676–682.

Goldsmith, H. H., Buss, A. H., Plomin, R., Rothbart, M. K., Thomas, A., Chess, S., et al. (1987). Roundtable: What is temperament? Four approaches. *Child Development, 58,* 505–529.

Graham, Y. P., Heim, C., Goodman, S. H., Miller, A. H., & Nemeroff, C. B. (1999). The effects of neonatal stress on brain development: Implications for psychopathology. *Development & Psychopathology, 11,* 545–565.

Gray, D., Woodward, L. J., Spencer, C., Inder, T. E., & Austin, N. C. (2006). Health service utilisation of a regional cohort of very preterm infants over the first 2 years of life. *Journal of Paediatrics & Child Health, 42,* 377–383.

Grunau, R. E., Weinberg, J., & Whitfield, M. F. (2004). Neonatal procedural pain and preterm infant cortisol response to novelty at 8 months. *Pediatrics, 114,* e77–e84.

Hack, M., Flannery, D. J., Schluchter, M., Cartar, L., Borawski, E., & Klein, N. (2002). Outcomes in young adulthood for very-low-birth-weight infants. *New England Journal of Medicine, 346,* 149–157.

Hack, M., Taylor, H. G., Klein, N., Eiben, R., Schatschneider, C., & Mercuri-Minich, N. (1994). School-age outcomes in children with birth weights under 750 g. *New England Journal of Medicine, 331,* 753–759.

Hanley, J. A., Negassa, A., Edwardes, M. D., & Forrester, J. E. (2003). Statistical analysis of correlated data using generalized estimating equations: An orientation. *American Journal of Epidemiology, 157,* 364–375.

Herrmann, C. (1997). International experiences with the Hospital Anxiety and Depression Scale—A review of validation data and clinical results. *Journal of Psychosomatic Research, 42,* 17–41.

Horwood, L. J., Mogridge, N., & Darlow, B. A. (1998). Cognitive, educational, and behavioural outcomes at 7 to 8 years in a national very low birthweight cohort. *Archives of Disease in Childhood Fetal & Neonatal Edition, 79,* F12–F20.

Hughes, M. B., Shults, J., McGrath, J., & Medoff-Cooper, B. (2002). Temperament characteristics of premature infants in the first year of life. *Journal of Developmental & Behavioral Pediatrics, 23,* 430–435.

Huppi, P. S., Murphy, B., Maier, S. E., Zientara, G. P., Inder, T. E., Barnes, P. D., et al. (2001). Microstructural brain development after perinatal cerebral white matter injury assessed by diffusion tensor magnetic resonance imaging. *Pediatrics, 107,* 455–460.

Inder, T. E., Anderson, N. J., Spencer, C., Wells, S., & Volpe, J. J. (2003). White matter injury in the premature infant: A comparison between serial cranial sonographic and MR findings at term. *American Journal of Neuroradiology, 24,* 805–809.

Inder, T. E., Warfield, S. K., Wang, H., Huppi, P. S., & Volpe, J. J. (2005). Abnormal cerebral structure is present at term in premature infants. *Pediatrics, 115,* 286–294.

Indredavik, M. S., Vik, T., Heyerdahl, S., Kulseng, S., & Brubakk, A. M. (2005). Psychiatric symptoms in low birth weight adolescents, assessed by screening questionnaires. *European Child & Adolescent Psychiatry, 14,* 226–236.

Katz, K. S., Dubowitz, L. M., Henderson, S., Jongmans, M., Kay, G. G., Nolte, C. A., et al. (1996). Effect of cerebral lesions on continuous performance test responses of school age children born prematurely. *Journal of Pediatric Psychology, 21,* 841–855.

Kersting, A., Dorsch, M., Wesselmann, U., Ludorff, K., Witthaut, J., Ohrmann, P., et al. (2004). Maternal posttraumatic stress response after the birth of a very low-birth-weight infant. *Journal of Psychosomatic Research, 57,* 473–476.

Klebanov, P. K., Brooks-Gunn, J., & McCormick, M. C. (1994). Classroom behavior of very low birth weight elementary school children. *Pediatrics, 94,* 700–708.

Kleberg, A., Westrup, B., & Stjernqvist, K. (2000). Developmental outcome, child behaviour and mother-child interaction at 3 years of age following Newborn Individualized Developmental Care and Intervention Program (NIDCAP) intervention. *Early Human Development*, *60*, 123–135.

Kopp, C. B. (1982). Antecedents of self-regulation—A developmental perspective. *Developmental Psychology*, *18*, 199–214.

Korkman, M., Liikanen, A., & Fellman, V. (1996). Neuropsychological consequences of very low birth weight and asphyxia at term: Follow-up until school-age. *Journal of Clinical & Experimental Neuropsychology*, *18*, 220–233.

Landry, S. H., & Chapieski, M. L. (1988). Visual attention during toy exploration in preterm infants: Effects of medical risk and maternal interactions. *Infants Behavior and Development*, *11*, 187–204.

Landry, S. H., Garner, P. W., & Swank, P. R. (1996). Effects of maternal scaffolding during joint toy play with preterm and full term infants. *Merrill-Palmer Quarterly*, *42*, 177–198.

Landry, S. H., Smith, K. E., & Swank, P. R. (2003). The importance of parenting during early childhood for school-age development. *Developmental Neuropsychology*, *24*, 559–591.

Landry, S. H., Smith, K. E., & Swank, P. R. (2006). Responsive parenting: Establishing early foundations for social, communication, and independent problem-solving skills. *Developmental Psychology*, *42*, 627–642.

Lawson, K. R., & Ruff, H. A. (2004). Early attention and negative emotionality predict later cognitive and behavioural function. *Internal Journal of Behavioral Development*, *28*, 157–165.

Lengua, L. J. (2003). Associations among emotionality, self-regulation, adjustment problems and positive mood in middle childhood. *Applied Developmental Psychology*, *24*, 595–618.

Lowe, J., Woodward, B., & Papile, L. A. (2005). Emotional regulation and its impact on development in extremely low birth weight infants. *Journal of Developmental & Behavioral Pediatrics*, *26*, 209–213.

Marlow, N. (2004). Outcome following extremely preterm birth. *Current Paediatrics*, *14*, 275–283.

Matas, L., Arend, R. A., & Sroufe, L. A. (1978). Continuity of adaptation in 2nd year—Relationship between quality of attachment and later competence. *Child Development*, *49*, 547–556.

McCormick, M. C., Workman-Daniels, K., & Brooks-Gunn, J. (1996). The behavioral and emotional well-being of school-age children with different birth weights. *Pediatrics*, *97*, 18–25.

McGrath, M. M., Sullivan, M., Devin, J., Fontes-Murphy, M., Barcelos, S., DePalma, J. L., et al. (2005). Early precursors of low attention and hyperactivity in a preterm sample at age four. *Issues in Comprehensive Pediatric Nursing*, *28*, 1–15.

Miles, M. S., Holditch-Davis, D., Schwartz, T. A., & Scher, M. (2007). Depressive symptoms in mothers of prematurely born infants. *Journal of Developmental & Behavioral Pediatrics*, *28*, 36–44.

Mouradian, L. E., Als, H., & Coster, W. J. (2000). Neurobehavioral functioning of healthy preterm infants of varying gestational ages. *Journal of Developmental & Behavioral Pediatrics*, *21*, 408–416.

Murray, L., & Cooper, P. (1997). Effects of postnatal depression on infant development. *Archives of Disease in Childhood*, *77*, 99–101.

Mykletun, A., Stordal, E., & Dahl, A. A. (2001). Hospital Anxiety and Depression (HADS) scale: Factor structure, item analyses and internal consistency in a large population. *British Journal of Psychiatry*, *179*, 540–544.

Nadeau, L., Boivin, M., Tessier, R., Lefebvre, F., & Robaey, P. (2001). Mediators of behavioral problems in 7-year-old children born after 24 to 28 weeks of gestation. *Journal of Developmental & Behavioral Pediatrics*, *22*, 1–10.

Nadeau, L., Tessier, R., Boivin, M., Lefebvre, F., & Robaey, P. (2003). Extremely premature and very low birthweight infants: A double hazard population? *Social Development*, *12*, 235–248.

National Institute of Child Health and Human Development Early Child Care Research Network. (1999). Child care and mother-child interaction in the first 3 years of life. *Developmental Psychology*, *35*, 1399–1413.

O'Connor, T. G., Heron, J., Glover, V., & Alspac Study. (2002). Antenatal anxiety predicts child behavioral/emotional problems independently of postnatal depression. *Journal of the American Academy of Child & Adolescent Psychiatry*, *41*, 1470–1477.

Ohgi, S., Takahashi, T., Nugent, J. K., Arisawa, K., & Akiyama, T. (2003). Neonatal behavioral characteristics and later behavioral problems. *Clinical Pediatrics*, *42*, 679–686.

Peterson, B. S., Anderson, A. W., Ehrenkranz, R., Staib, L. H., Tageldin, M., Colson, E., et al. (2003). Regional brain volumes and their later neurodevelopmental correlates in term and preterm infants. *Pediatrics*, *111*, 939–948.

Porges, S. W. (1996). Physiological regulation in high risk infants: A model for assessment and potential intervention. *Development and Psychopathology*, *8*, 43–58.

Posner, M. I., & Rothbart, M. K. (2000). Developing mechanisms of self-regulation. *Development and Psychopathology*, *12*, 427–441.

Rose, S., Feldman, J. F., Wallace, I. F., & McCarton, C. (1989). Infant visual attention: Relation to birth status and developmental outcome during the first five years. *Developmental Psychology*, *25*, 560–576.

Segal, L. B., Oster, H., Cohen, M., Caspi, B., Myers, M., & Brown, D. (1995). Smiling and fussing in seven-month-old preterm and full-term black infants in the still-face situation. *Child Development*, *66*, 1829–1843.

Shaw, D. S., Keenan, K., Vondra, J. I., Delliquadri, E., & Giovannelli, J. (1997). Antecedents of preschool children's internalizing problems: A longitudinal study of low-income families. *Journal of the American Academy of Child & Adolescent Psychiatry, 36*, 1760–1767.

Shields, A., & Cicchetti, D. (1997). Emotion regulation among school-age children: The development and validation of a new criterion Q-sort scale. *Developmental Psychology, 33*, 906–916.

Silk, J. S., Shaw, D. S., Skuban, E. M., Oland, A. A., & Kovacs, M. (2006). Emotion regulation strategies in offspring of childhood-onset depressed mothers. *Journal of Child Psychology & Psychiatry & Allied Disciplines, 47*, 69–78.

Statistics New Zealand. (2001). *Census snapshot: Maori.* Retrieved July 21, 2008, from http://www.stats.govt.nz/products-and-services/Articles/census-snpsht-maori-Apr02.html.

Stewart, A. L., Rifkin, L., Amess, P. N., Kirkbride, V., Townsend, J. P., Miller, D. H., et al. (1999). Brain structure and neurocognitive and behavioural function in adolescents who were born very preterm. *Lancet, 353*, 1653–1657.

Supplee, L. H., Shaw, D. S., Hailstones, K., & Hartman, K. (2004). Family and child influences on early academic and emotion regulatory behaviors. *Journal of School Psychology, 42*, 221–242.

Sykes, D. H., Hoy, E. A., Bill, J. M., McClure, B. G., Halliday, H. L., & Reid, M. M. (1997). Behavioural adjustment in school of very low birthweight children. *Journal of Child Psychology & Psychiatry & Allied Disciplines, 38*, 315–325.

Taylor, H. G., Minich, N., Bangert, B., Filipek, P. A., & Hack, M. (2004). Long-term neuropsychological outcomes of very low birth weight: Associations with early risks for periventricular brain insults. *Journal of the International Neuropsychological Society, 10*, 987–1004.

Thompson, R. A. (1994). Emotion regulation: A theme in search of definition. *Monographs of the Society for Research in Child Development, 59*, 25–52.

Vicari, S., Caravale, B., Carlesimo, G. A., Casadei, A. M., & Allemand, F. (2004). Spatial working memory deficits in children at ages 3-4 who were low birth weight, preterm infants. *Neuropsychology, 18*, 673–678.

Wechsler, D. (1989). *WPPSI-R manual: Wechsler preschool and primary scale of intelligence*. New York: Psychological Corporation.

Woodward, L. J., Anderson, P. J., Austin, N. C., Howard, K., & Inder, T. E. (2006). Neonatal MRI to predict neurodevelopmental outcomes in preterm infants. *New England Journal of Medicine, 355*, 685–694.

Woodward, L. J., Edgin, J. O., Thompson, D., & Inder, T. E. (2005). Object working memory deficits predicted by early brain injury and development in the preterm infant. *Brain, 128*, 2578–2587.

Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavia, 67*, 361–370.

Child Development, September/October 2008, Volume 79, Number 5, Pages 1463–1476

Neighborhood Poverty and Early Transition to Sexual Activity in Young Adolescents: A Developmental Ecological Approach

Véronique Dupéré and Éric Lacourse
Université de Montréal

J. Douglas Willms
University of New Brunswick

Tama Leventhal
Tufts University

Richard E. Tremblay
Université de Montréal

This study examined how the link between neighborhood poverty and the timing of sexual initiation varies as a function of age, gender, and background characteristics. A sample of $N = 2,596$ predominately White Canadian adolescents from the National Longitudinal Survey of Children and Youth was used. Sexual initiations occurring between 12 and 15 years old were considered. Results showed that younger adolescent females who lived in poor neighborhoods and who had a history of conduct problems were more likely to report early sexual activity. Peer characteristics partly accounted for this susceptibility. Among adolescent males, no direct neighborhood effects were found, but those who had combined risks at multiple levels appeared more vulnerable. The theoretical and practical implications of these findings are discussed.

For many adolescents, sexual experimentation includes navigating the transition to first heterosexual intercourse, an event that becomes relatively common by the end of adolescence (Darroch, Frost, Singh, & Study Team, 2001). Although this transition becomes normative by middle to late adolescence, a small group of adolescents experience it much sooner. In particular, about 5% have had sex at 12 years of age in the United States; this rate gradually rises to around 10% at age 13 and 19% at age 14 (Albert, Brown, & Flanigan, 2003). Compared with later initiators, early initiators are much more likely to contract sexually transmitted diseases, undergo unwanted teen pregnancy, and report involuntary sexual experiences (e.g., Albert et al., 2003; Kaestle, Halpern, Miller, & Ford, 2005). Reports of such negative experiences decrease sharply for those transitioning in later adolescence. Despite the significance of sexual initiation in early adolescence, a recent review noted that its determinants have been understudied (Albert et al., 2003). This is problematic because the processes at play in older adolescents do not automatically apply to their younger peers. Just as the consequences of sexual activity dramatically change as adolescents mature, developmentally oriented researchers proposed that its determinants must also fluctuate at different developmental stages (Whitbeck, Yoder, Hoyt, & Conger, 1999).

In this study, we propose that both developmental (i.e., age) and contextual (i.e., neighborhood and peers) factors need to be considered simultaneously to adequately understand the determinants of the timing of sexual initiation. Adolescence is a period of intensified participation in larger social contexts outside the home, such as neighborhoods and peer groups, and these contexts are thought to exert powerful influences upon adolescent development, including sexuality (Browning, Leventhal, & Brooks-Gunn, 2004). The current study adopts an ecological, developmental approach to examine the timing of young adolescents' sexual initiation. That is, longitudinal data are used to explore how the relationship between neighborhood poverty and sexual initiation and hypothesized mediators (i.e., peer characteristics) vary with age among adolescent males and females. In addition, the moderating role of neighborhood circumstances on family and individual risk factors is examined.

This study was based on the doctoral dissertation of the first author, who was supported by a scholarship from Social Sciences and Humanities Research Council of Canada. The authors wish to thank Eric Dion and anonymous reviewers for comments on previous drafts. The research and analysis are based on data from Statistics Canada, and the opinions expressed do not necessarily represent the views of Statistics Canada.

Correspondence concerning this article should be addressed to Eric Lacourse, Département de Sociologie, Université de Montreal, CP 6128, succursale Centre-Ville, Montréal, Québec, Canada H3C-3J7. Electronic mail may be sent to eric.lacourse@umontreal.ca.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7905-0016

Neighborhood Disadvantage and Early Transitions to Sexual Activity: Potential Moderators

Ecological models of human development suggest that the neighborhood environment is an important context for development (Bronfenbrenner, 1979). Correspondingly, studies have repeatedly found a link between neighborhood disadvantage and a number of risky adolescent sexual outcomes, such as inconsistent contraceptive use or teenage childbearing (Baumer & South, 2001; Leventhal & Brooks-Gunn, 2000). In contrast, studies investigating the link between neighborhood disadvantage and *timing* of sexual initiation have produced mixed results. Some found independent effects of neighborhood structural characteristics (i.e., poverty) over and above other important determinants (e.g., Browning et al., 2004; Ramirez-Valles, Zimmerman, & Juarez, 2002), whereas others did not (e.g., Baumer & South, 2001; Upchurch, Aneshensel, Sucoff, & Levy-Storms, 1999). Sample differences in terms of participants' characteristics, notably age and gender, might explain these inconsistencies.

Theoretical and empirical considerations suggest that the association between neighborhood disadvantage and sexual initiation could vary as a function of age. The developmental model proposed by Whitbeck et al. (1999) argues that exposure to nonnormative social contexts, such as deviant peers, should predict transitions occurring in early adolescence when the transition is clearly nonnormative and gradually lose predictive power as adolescents mature and the transition becomes more normative. Although they did not explicitly discuss neighborhood factors, their reasoning suggests that neighborhood disadvantage could be associated with sexual initiation in early adolescence, but not necessarily at later ages, given that neighborhood effects are often interpreted as a consequence of exposure to neighborhood deviant peers (Baumer & South, 2001). Also, younger teens may be especially susceptible to peer pressure and peer norms (Brooks-Gunn & Furstenberg, 1989). Empirical evidence indirectly supports the hypothesis of age-graded neighborhood effects on sexual initiation because significant effects have tended to be observed in comparatively younger samples. For instance, Browning et al. (2004) obtained significant neighborhood effects when modeling transitions occurring between 11 and 16 years old, whereas others did not, in samples followed throughout adolescence and into early adulthood (Baumer & South, 2001).

Gender is also likely to moderate the association between neighborhood disadvantage and early sexual activity. If sexual activity represents a departure from the norm for early adolescents, it is especially the case for females. Norms tend to recommend chastity for adolescent females, whereas sexual experimentation is often tacitly encouraged in males (Benda & DiBlasio, 1994). Echoing these diverging norms, sexually active adolescent females are more likely to be labeled negatively by their peers and to express remorse and regrets regarding their early sexual experiences, as compared with adolescent males (Dickson, Paul, Herbison, & Silva, 1998; Graber, Brooks-Gunn, & Galen, 1998). Because social norms tend to disapprove early sexual activity more strongly among adolescent females, exposure to more permissive norms might be especially influential for them. Empirical results suggest that in disadvantaged neighborhoods, young adolescent females are more likely to be exposed to alternative cultural scripts regarding sexuality (Harding, 2007). For instance, Harding (2007) found that in disadvantaged neighborhoods, a sizable minority of adolescents tend to view adolescent pregnancy in a positive light, whereas in more advantaged neighborhoods, there is a general consensus to view adolescent pregnancy as a negative event. Adolescent females appear especially sensitive when exposed to such alternative scripts. Indeed, exposure to peers having more permissive attitudes and behaviors toward sexuality has been found to predict entry into sexual activity for adolescent females but not for adolescent males (Billy & Udry, 1985). For this reason, adolescent females' sexual behavior could be generally more susceptible to neighborhood effects.

On average, adolescents living in poor neighborhoods are at an increased risk of being exposed to alternative cultural scripts regarding sexuality (Harding, 2007), but there is certainly a great deal of individual variability in the degree and intensity of exposure. Mounting evidence shows that key family and individual characteristics modulate neighborhood effects. At the family level, low socioeconomic status (SES), family disruption, and residential instability are factors likely to increase exposure to neighborhood peers with more permissive attitudes. For instance, single parents tend to have fewer resources to exert effective supervision than two-parent families and inconsistent supervision is more strongly associated with early sexual activity for adolescents living in disadvantaged neighborhoods (Browning, Leventhal, & Brooks-Gunn, 2005). Also, residentially unstable adolescents are at risk of making new deviant friends when they move into a new neighborhood and this, in turn, increases the likelihood of early initiation (South, Haynie, & Bose, 2005). This association could be especially strong among adolescents moving into a disadvantaged neighborhood. At the individual level, neighborhood disadvantage might also exacerbate the accelerating impact of early pubertal development and conduct problems, two characteristics associated with early sexual activity (Browning et al., 2004). Indeed, neighborhood disadvantage increases the chances that such predisposing individual characteristics actually translate into problem behaviors (e.g., Obeidallah, Brennan, Brooks-Gunn, & Earls, 2004). Thus, there are good reasons to consider family and individual characteristics not only as statistical controls but also as potential moderators of neighborhood effects.

Neighborhood Poverty and Early Sexual activity: A Peer Mediation Model

So far, we have argued that neighborhood disadvantage may be associated with early sexual activity and that this association may vary as a function of individual and family characteristics. This section now addresses potential explanations for neighborhood effects. Theoretical and empirical considerations suggest that "peers may be the primary agent through which community socialization adversely affects adolescents" (Leventhal & Brooks-Gunn, 2000, p. 327). Previous research suggests that adolescents are more likely to affiliate with deviant peers if they live in a disadvantaged neighborhood (Brody et al., 2001) apparently because of a reduced collective capacity to supervise and control youth groups in these contexts (Sampson, Morenoff, & Gannon-Rowley, 2002). The proposition that peer processes are important mediators of neighborhood effects has been repeatedly supported in the delinquency literature (e.g., Ingoldsby et al., 2006; Simons, Johnson, Beaman, Conger, & Whitbeck, 1996).

Recent findings suggest that peer factors may play a primary mediating role for sexual risk taking as well. When examining the mediating role of a host of potential explanatory processes for neighborhood effects on adolescents' sexual behaviors, Baumer and South (2001) found that the sole peer characteristic that they considered, a composite measure of peer attitudes and behaviors, was the only variable with a significant, although partial, mediating effect. The present study builds on these findings by considering the mediating role of peer deviance and of involvement with an older sexual partner. These characteristics were selected because evidence suggests that they are associated with both neighborhood disadvantage and early sexual activity. We briefly outline this evidence below.

Affiliation with deviant peers is one of the prominent risk factors for early sexual initiation (e.g., French & Dishion, 2003), most likely because deviant peers provide a pool of willing partners and cultivate a sense that early sexual activity is desirable. Deviant peer affiliation could mediate neighborhood effects, given that neighborhood disadvantage is a risk factor for deviant peer affiliation (e.g., Brody et al., 2001). Involvement with an older partner is another potential explanatory factor, especially for adolescent females. Indeed, although getting involved with an older partner increases the risk of early sexual activity for both adolescent males and females (Marin, Coyle, Gomez, Carvajal, & Kirby, 2000), this situation is rare among adolescent males and affects a much greater proportion of females (Halpern, Kaestle, & Hallfors, 2007), especially young, early-maturing females (Manlove, Ryan, & Franzetta, 2007). This may be attributable to the attractiveness of young, physically developed adolescent females to older males (Caspi, Lynam, Moffitt, & Silva, 1993; Halpern et al., 2007). Certain contexts, such as disadvantaged neighborhoods, may increase the likelihood that physically mature adolescent females affiliate with older peers and engage in risky behaviors (Obeidallah et al., 2004). Thus, differential exposure to older partners and deviant peers could potentially explain neighborhood effects on the timing of sexual initiation.

In this study, a developmental, ecological perspective is adopted to examine young adolescents' transitions to first sexual activity. A representative sample of Canadian adolescents is used to explore how the relationship between ecological risk factors (neighborhood disadvantage), hypothesized mediators (peer characteristics), and sexual initiation varies with age. In addition, the moderating effects of other factors that could alter adolescent susceptibility to neighborhood circumstances, such as gender and individual characteristics, are considered. We hypothesize that peer characteristics would at least partly account for this vulnerability.

Method

Sample

The National Longitudinal Survey of Children and Youth (NLSCY) is a nationally representative, prospective study of Canadian children and adolescents (Statistics Canada, 2005). The sample design is a clustered probability sample of private households within the 10 Canadian provinces, excluding children living in remote areas, institutional settings, and on First Nations reserves. The NLSCY was launched in 1994–1995, with follow-up surveys conducted every 2 years thereafter. Initially, 13,439 households of 15,579 identified as having at least one dependent child, ages newborn to 11 years, agreed to participate in Cycle 1 of the study. Within each household, up to two children per family were selected to participate in the longitudinal survey, resulting in a longitudinal sample of 16,903. In each of the survey households, the "person most knowledgeable" about the child was interviewed. In virtually all cases, this person was a parent and in most cases the mother and hereafter is referred to as such. From 10 years old on, children were also asked to complete self-report questionnaires.

This study uses data from the first five data collection cycles (1994–1995 to 2002–2003). Because questions about age at sexual initiation were consistently included in the survey only for participants of at least 16 years old, the subsample used includes older NLSCY participants who were at least 16 years old in Cycle 5 (i.e., participants who were between 8 and 11 years old in Cycle 1, in 1994–1995). This represents about 5,000 adolescents. Of these, only those who participated in all cycles of data collection were selected. A total of 2,960 participants met this inclusion criterion, representing a retention rate of about 60% for an 8-year span. Of that number, 19 cases who reported sexual initiation before age 12 were excluded. Another 23 participants were excluded due to data inconsistencies. Finally, one child per household was selected at random to avoid within-family clustering. Thus, the final sample size is $N = 2,596$. The sample was evenly distributed across both genders and more than 90% of the participants were Whites. The remaining participants were composed mainly of First Nations children and children of Asian or African descent.

To take attrition into account as well as the stratified sampling of the NLSCY, all analyses were conducted using normalized longitudinal survey weights. The use of weights helps preserve the representativeness of the sample despite survey dropout, but it does not eliminate potential bias associated with partial nonresponse. Partial nonresponse occurs when participants fail to answer some of the survey questions. In this study, the rate of partial nonresponse varied between 0% and 7% for parent-reported variables and between 3% and 15% for self-reported variables (by design, some self-reported variables had a higher rate of missing data; for instance, those who were still virgins in Cycle 5—i.e., at 16–17/18–19 years old—were not asked about the age of first sexual partner). Multiple imputation was used to handle partial nonresponse (Allison, 2001). Multiple imputation is among the best methods currently available for reducing possible bias due to partial nonresponse, whereas taking into account imputation uncertainty. In this study, partial nonresponse was treated through the creation of five complete data sets obtained through multiple imputation (SAS MI procedure). Regression analyses were performed on each imputed data set separately, and the parameter estimates obtained in each set were recombined, resulting in a final, single set of estimates. The analyses were also repeated using simple listwise deletion of cases with missing data, and the main conclusions of the present study remained essentially unchanged.

Measures

Transitions to first sexual activity occurring during the first half of adolescence were considered (12–15 years old). Family (family structure, SES, and residential instability) and individual (behavior problems and pubertal development) predisposing factors were measured in late childhood (10–11 years old) and were thus antecedent to the outcome. Neighborhood poverty was measured concurrently during early adolescence (12–15 years old). Peer deviance was also measured concurrently, at 12–13 and 14–15 years old, whereas age at sexual intercourse and age of first sexual partner were measured retrospectively (16–17 or 18–19 years old). A summary of the measurement time points for each variable is provided in Table A1 in the Appendix.

Late Childhood Family and Individual Risk Factors

Family SES was measured through an index computed by Statistics Canada (Willms & Shields, 1996). Five standardized variables were combined to calculate family SES, including mothers' and fathers' educational attainment, mothers' and fathers' occupational prestige (classified according to the Standard Occupational Classification System), and household income. In this sample, the index ranged from −3.5 to 2.8 ($M = 0.0$, $SD = 0.8$). *Intact family status* represented those (73%) who lived with two biological or adoptive parents. *Residential instability during childhood* was also considered. The number of childhood relocations before the child reached age 10–11 varied from 0 to 15. This variable was recoded into a dichotomous variable contrasting those who had moved fewer than five times with those who had moved five times or more (11%).

Conduct problems were assessed with nine self-reported items representative of the three symptom domains listed in the *Diagnostic and Statistical Manual of Mental Disorders, 4th edition (DSM-IV)* for conduct disorders, that is, physical aggression (bullying, fighting, attacking, kicking, or biting), destruction of property, deceitfulness or theft (vandalism and stealing), and serious violations of rules (running away, truancy, and staying out all night). Following the *DSM-IV* classification rules, those who reported that they had engaged in at least three of these behaviors in the past 12 months (13%) were considered at risk. *Early pubertal development* was measured with three gender-specific items pertaining to change in physical development markers: body hair, breast development and menstruation for adolescent females, and body hair, facial hair, and voice change for adolescent males. All questions were answered on a 4-point scale (0 = *no change yet* to 3 = *change completed*), except for menstruation, which was dichotomous. For items measured on a 4-point scale, responses were rescaled to range from 0 to 1. Then, total scores were computed and normalized by gender and age. Those with higher scores were more physically developed than their same-gender peers in late childhood.

Neighborhood Poverty

Links to the 2001 Canadian Census were made to assess the neighborhood poverty status associated with participants' main place of residence. The dissemination area (DA) was the geographic unit used to approximate the neighborhood environment. The DA has a population range of about 400 – 700 individuals (Puderer, 2001), which roughly corresponds to U.S. Census block groups. All DAs with 20% or more residents under Statistics Canada's low-income cutoff were considered as poor neighborhoods. About a quarter of participants resided in a poor neighborhood.

Peer Characteristics

Peer deviance was estimated with a single item asking the mother how often their child "hung around with kids who are frequently in trouble." Those who exhibited this behavior often or sometimes were considered at risk (12%). This single item was retained because unfortunately, other items measuring peer deviance were not consistent across data collection cycles. Participants who were sexually experienced were asked retrospectively about the *age of their first sexual partner*. For those who answered this question in more than one cycle, the first valid age reported was used. In virtually all instances, reported age of partners ranged between 12 and 25 years old. A dichotomous variable distinguishing those who had a first partner at least 3 years older than themselves at time of sexual initiation was created. As expected, adolescent females were much more likely to have been sexually initiated with an older partner as compared to adolescent males, especially among younger adolescents. For instance, among adolescent females who became sexually initiated before 16 years old, 34% reported first intercourse with an older partner, whereas this rate dropped to 4% for adolescent males. These proportions are similar to those of other studies showing that around a third of sexually initiated adolescent females have a first sexual experience with a partner at least 3 years older, whereas this proportion is much lower among males (Manlove et al., 2007).

Outcome: Timing of First Consensual Sexual Intercourse

The observation window for the outcome covered the first half of adolescence (12 – 15 years old). The timing of entry into sexual activity was measured retrospectively by asking participants: "Have you ever had consensual sexual intercourse?" And, for those who answered positively: "How old were you when you first had consensual sexual intercourse?" For those who answered the question about age at sexual initiation in more than one cycle, the first valid age reported was used. Otherwise, the answer obtained in the last available wave of data collection was used (Cycle 5). It is important to note that sexual intercourse was not explicitly defined in the questionnaire. Thus, the measure is subject to interpretation bias and does not allow distinctions between sexual initiations occurring between same- and other-gender partners. The distribution of age at first intercourse was not significantly different for adolescent males and females, $\chi^2(4, N = 2{,}596) = 2.97$, $p = 0.56$, a result consistent with other recent studies using representative samples of Canadian adolescents (Hansen, Mann, McMahon, & Wong, 2004). As expected, the proportion of adolescents becoming sexually initiated increased steadily with age, from less than 2% at 12 years old to up to 15% at 15 years old. These proportions are consistent with those obtained in other representative samples of Canadian adolescents (Hansen et al., 2004; Maticka-Tyndale, McKay, & Barrett, 2001).

Results

Descriptive Statistics

Table 1 shows the distribution of the risk factors as a function of neighborhood poverty. Adolescents living in poor neighborhoods were more likely to come from low-SES, nonintact, and residentially

Table 1

Descriptive Statistics as a Function of Neighborhood Context[a]

	Nonpoor neighborhood (n = 1,948)			Poor neighborhood (n = 648)			
	%	*M*	*SD*	%	*M*	*SD*	χ^2/F
Late childhood risk factors							
Family socioeconomic status		0.14	0.74		−0.26	0.83	158.5***
Intact family	78.3			57.3			126.7***
Residentially unstable family	9.3			14.3			14.9***
Conduct problems	12.3			13.4			0.5
Pubertal development		−0.06	0.99		0.01	1.02	3.3†
Peer characteristics (age in years)							
Deviant peers (12–13)	8.2			13.5			18.4***
Deviant peers (14–15)	12.6			16.1			5.9*
Older partner (12–13)[b]	34.1			46.2			2.3
Older partner (14–15)[b]	15.9			11.9			1.7
Transitions to sexual activity							
12 years old	1.3			3.5			14.9***
13 years old	2.9			4.9			6.8**
14 years old	6.8			7.9			1.1
15 years old	14.6			17.0			2.7

[a] Rounded data from first imputed data set (rounded down to closest lower integer for age at sexual initiation). [b] For those who transitioned to first sexual activity at 12–13 or 14–15 years old.

†$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$.

unstable families. However, no significant differences were found among the individual predisposing factors (conduct problems and pubertal development). Results also show that in poor neighborhoods, adolescents were more likely to affiliate with deviant peers. Finally, as expected, the age distribution for transitions to first sexual activity was significantly different in poor and more affluent neighborhoods, with significantly higher rates of transitions at 12 and 13 years old in poor neighborhoods.

Discrete-Time Hazard Models

To estimate the association between the risk factors and the timing of first sexual activity, discrete-time hazard models were used (Singer & Willett, 2003). This strategy is appropriate because the timing of first sexual activity was measured in terms of years, a discrete-time scale. Regular logistic regression routines can be used to estimate this type of model. In this case, the SAS LOGISTIC procedure estimated the regression coefficients separately in each of the five imputed data sets. Then, the MIANALYSE procedure recombined the results to obtain a final, single set of parameter estimates (Allison, 2001). Also, interactions between all risk factors and linear time were included. To test for these interactions, time was coded as a continuous variable ranging from 0 = *12 years old* to 3 = *15 years old*.

Bivariate Results

Table 2 shows results from bivariate discrete-time hazard models performed for adolescent females and males. For each gender, the first column presents the estimated odds ratio (OR) at 12 years old. For instance, the estimated OR associated with low family SES is 2.55 at 12 years old for adolescent females. This means that the odds of transitioning to sexual activity at 12 years old are about 2.5 times higher for adolescent females from low-SES families (i.e., with an SES score at least 1 *SD* below the national mean), compared with adolescent females from higher SES families. The second column contains the interaction effects with time, that is, the attenuating effect of every passing year. Continuing with the SES example, the first result in the second column for adolescent females indicates that the value of the OR associated with low SES diminishes by 24% ($1.00 - 0.76 = 0.24$) with every passing year. At 13 years old, the estimated OR becomes 1.94 ($2.55 \times 0.76 = 1.94$) and, finally, at 15 years old, it is 1.12 ($2.55 \times 0.76^3 = 1.12$). Thus, the effect of low family SES is strong in the early years of adolescence, but gradually fades away as adolescent females grow older, becoming negligible at 15 years old.

For adolescent females, the bivariate results show a significant effects for all the variables, except early pubertal development (although when the interaction

Table 2

Bivariate Discrete-Time Hazard Models Including Interaction Effects With Time: Estimated OR at 12 Years Old and OR Representing the Attenuating Effect of Time[a]

	Adolescent females (n = 1,322)		Adolescent males (n = 1,274)	
	Initial OR (12 years old)	Time OR[b]	Initial OR (12 years old)	Time OR[b]
Late childhood risk factors				
Low family SES[c]	2.55***	0.76***	1.70***	0.91
Intact family	0.27***	1.32*	0.29***	1.17
Family residential instability	6.05***	0.62***	3.60***	0.71[†]
Conduct problems	6.49***	0.55***	5.00***	0.60*
Pubertal development[d]	1.03	1.02	1.55*	0.87
Neighborhood context				
Neighborhood poverty	4.81***	0.55***	1.03	1.05
Peer characteristics				
Deviant peers	14.59***	0.64**	2.32	0.94
Older partner	28.50***	0.30***	10.49**	0.46

Note. OR = odds ratio; SES = socioeconomic status.

[a] Dummy markers of age were included, but associated intercepts are not shown. [b] ORs in this column represent the interaction effects with time. For instance, the OR for family SES at 12 years old for adolescent females is 2.55. At 13 years old, it becomes 1.94 (2.55 × 0.76 = 1.94) and so on. [c] One standard deviation below mean. [d] One standard deviation above mean.

[†] $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

effect with time was excluded, the OR for this variable approached significance; OR = 1.09, 95% CI = 0.99–1.21). The ORs are especially large for affiliation to deviant peers and involvement with older partners. Low family SES, nonintact family status, and family residential instability, along with conduct problems and neighborhood poverty, were associated with higher risks of early sexual activity. Results also show that the effects of these risk factors were stronger in early adolescence and gradually faded as adolescent females grew older. The absence of significant effect for pubertal development is consistent with other NLSCY studies reporting no direct association between pubertal development measured at 10–11 years old and early sexual activity or other problem behavior in early adolescence among girls (Pepler & Sedighdeilami, 1998; Shapka, Dahinten, & Arım, 2006), although evidence suggests that when measured at later ages, this variable has a significant impact on the likelihood of being sexually experienced for both males and females (Garriguet, 2005). Thus, this result ought to be interpreted while keeping in mind the restrained variance of this variable at 10–11 years old (see Dorn, Susman, & Ponirakis, 2003; Udry & Billy, 1987, for a discussion of how the timing of measurement influence the estimated impact of pubertal development on adolescent behavior).

For adolescent males, the effects of the risks factors were generally smaller than for females, especially for peer characteristics. As opposed to females, family and individual predispositions were more important determinants of adolescent males' sexual behavior than peer processes. In this regard, the results are consistent with previous results showing that peers are more influential with respect to adolescent females' sexual behavior (Billy & Udry, 1985). Also, the effect of the risk factors tended to be more evenly distributed across time. For adolescent males, only one significant interaction effect with time was found (conduct problems) along with one marginally significant interaction effect (residential instability).

Bivariate models including interaction effects with gender were also examined for the combined sample to determine whether the differences between adolescent males' and females' coefficients were statistically significant (results not shown). Significant or marginally significant interaction effects were found for SES (B = .50, SE = 0.28; p = .087), pubertal development (B = .41, SE = 0.20; p = .041), neighborhood poverty (B = −1.54, SE = 0.71; p = .055), and peer deviance (B = −1.84, SE = 0.60; p = .006).

Multivariate Results

Three multivariate models were estimated for adolescent males and females. They gradually incorporate the predictors, starting with the most distal ones. The first model includes late childhood individual and family control variables, along with neighborhood poverty. In the second model, interaction terms between neighborhood disadvantage and predisposing individual and family risk factors are added. Only the interaction terms that were significant in preliminary analyses are included. In a final step, the more proximal peer processes were incorporated. This last step allows us to examine the potential mediating effect of peer processes. In all three models, interactions with time are included when appropriate.

Adolescent females. Table 3 presents the multivariate results for adolescent females. Model 1 shows that those living in poor neighborhoods were more likely to transition early, even after accounting for individual and family predisposing factors. The ORs were especially high in early adolescence and gradually faded as adolescent females grew older. For instance, at 12 years old, the odds of transitioning to first sexual activity were about 2.5 times higher in poor neighborhoods, but at 15 years old, the estimated ORs decreased to 0.74 ($2.46 \times 0.67^3 = 0.74$).

Model 2 reveals a significant interaction effect between neighborhood poverty and conduct problems and a marginally significant effect with residential instability. Because the OR associated with neighborhood poverty is no longer significant, adolescent females living in poor neighborhoods were more vulnerable *only* if they had a history of conduct problems and marginally so if they came from a residentially unstable family. To facilitate the interpretation of these interaction effects, Figure 1 presents the predicted probabilities of transition to first sexual activity as a function of age for various configurations of risk factors. The figure shows that among adolescent females residing in poor neighborhoods, only those with a history of conduct problems were more likely to transition early. The figure also confirms that the largest effects are found in early adolescence and gradually decline as adolescent females grow older. For instance, among adolescent females with conduct problems, the predicted hazard of transitioning to first intercourse, while holding other variables at their average values, is about 7.1 times higher in disadvantaged neighborhoods at 12 years old but only 1.2 times at 15 years old.

Finally, the last model in Table 3 shows that when the peer processes are included in the equation, the size of the interaction effects previously discussed is significantly reduced (e.g., the ORs are reduced by about 40% ($[6.69-3.78]/6.69$) for the interaction between neighborhood disadvantage and conduct problems) and the interaction terms become nonsignificant, thus indicating that peer characteristics partially accounted for the interaction effects found between individual and neighborhood risks. Also consistent with the proposed peer mediation model, Figure 2 illustrates that the peer risk factors are much

Table 3

Multivariate Discrete-Time Hazard Models Including Interaction Effects With Time for Adolescent Females: Estimated OR at 12 Years Old and OR Representing the Attenuating Effect of Time[a]

	Model 1		Model 2		Model 3	
	Initial OR	Time OR	Initial OR	Time OR	Initial OR	Time OR
Late childhood risk factors						
Low SES[b]	1.71***	0.86**	1.51†	0.93	1.32	0.96
Intact family	0.76	0.94	0.87	0.87	0.84	0.91
Residential instability	2.97**	0.76†	1.38	1.00	0.84	1.14
Conduct problems	2.75**	0.73†	1.07	1.00	0.88	0.93
Pubertal development[c]	1.06		1.06		1.09†	
Neighborhood context						
Neighborhood poverty	2.46*	0.67**	1.63	0.74	2.18	0.68
Neighborhood Context × Childhood Risk[d]						
Neighborhood Poverty × Residential Instability			4.01†	0.63	2.64	0.73
Neighborhood Poverty × Conduct Problems			6.69*	0.59	3.78	0.86
Peer characteristics						
Deviant peers					6.23***	0.87
Older partner					15.33***	0.35***

Note. OR = odds ratio; SES = socioeconomic status.

[a] Dummy markers of age were included, but associated intercepts are not shown. [b] One standard deviation below mean. [c] One standard deviation above mean. [d] Interaction effects between neighborhood poverty and SES and intact family status were included in Models 2 and 3, but these interactions were nonsignificant in the multivariate models; therefore, associated ORs are not shown.

†$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$.

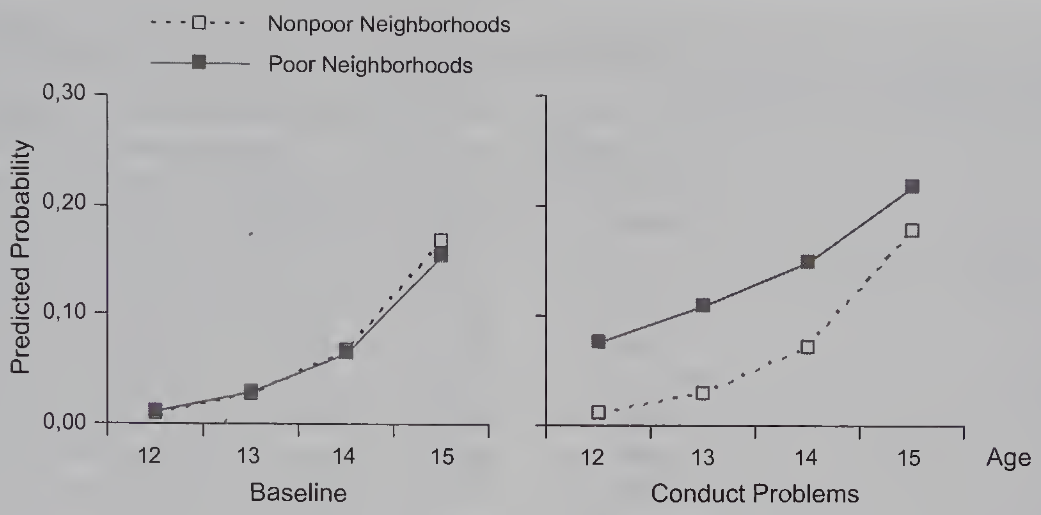


Figure 1. Adolescent females' predicted probabilities of transition to first sexual activity at baseline and among those with conduct problems in poor and nonpoor neighborhoods, as a function of age.

more prevalent for adolescent females with a history of conduct problems *and* who live in poor neighborhoods. For adolescent females in general, living in a poor neighborhood had little impact on peer characteristics.

Adolescent males. Table 4 presents the multivariate results for adolescent males. Consistent with the bivariate results, Model 1 shows that after controlling for preexisting family and individual risks, neighborhood disadvantage is not, on average, associated with accelerated transitions to first sexual activity. In contrast, nonintact family status, previous conduct problems, and early pubertal development all had independent, accelerating effects. Only one significant interaction effect with time emerged. This showed that the effect of previous conduct problems is stronger in younger adolescent males and attenuates as adolescent males grow older. Indeed, for those with a history of conduct problems, the estimated odds were 4.1 times higher at 12 years old but only 1.1 times higher at 15 years old ($4.1 \times 0.64^3 = 1.1$).

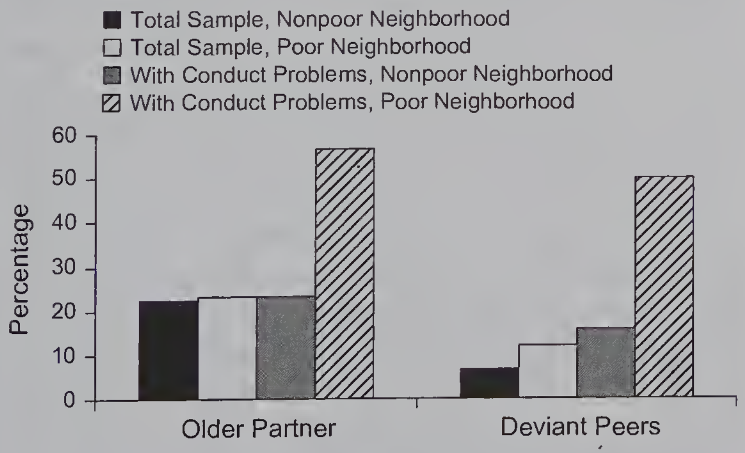


Figure 2. Adolescent females' peer characteristics as a function of neighborhood poverty status: Reported proportion of older partners (among those who transitioned to first sexual activity at 12 or 13 years old) and of deviant peers (at 12–13 years old) for adolescent females' total sample and among those with a history of conduct problems.

In Model 2, significant interaction effects between late childhood risk factors and neighborhood poverty emerged for conduct problems and family SES. Figure 3 unpacks these effects. It plots adolescent males' estimated probabilities of transition to first sexual activity as a function of age for different configurations of risk factors. Figure 3 shows that the differences are smaller compared with the results for females (Figure 2). In Figure 3, the only curve that is clearly distinguishable from the others is the one representing the probabilities of transition for adolescent males living in a poor neighborhood and who had histories of conduct problems and who were from low-SES families. As shown in the figure, the predicted probability of transitioning to first intercourse, while holding other variables at their average values, is about 3 times higher in disadvantaged neighborhoods at 12 years old and 1.8 times higher at 13 years old for adolescent males who combined risks at the neighborhood, family, and individual levels.

Model 3 examines whether these effects are attributable to the peer processes considered in this study. Results show that when the other variables are controlled, deviant peer affiliation and having an older partner are only marginally associated with early sexual activity and that these variables do not mediate the interaction effects discussed above. Indeed, the interaction effects remain significant or marginally significant and their magnitude is not significantly altered. In this final model for adolescent males, apart from the interaction effects discussed above, only the parameters associated with family structure and pubertal development remain significant.

Table 4
Multivariate Discrete-Time Hazard Models Including Interaction Effects With Time for Adolescent Males: Estimated OR at 12 Years Old and OR Representing the Attenuating Effect of Every Passing Year[a]

	Model 1		Model 2		Model 3	
	Initial OR	Time OR	Initial OR	Time OR	Initial OR	Time OR
Late childhood risk factors						
Low SES[b]	$1.47^{\dagger}$	0.92	1.18	1.04	1.12	1.03
Intact family	0.48**		0.46**		0.48**	
Residential instability	$2.31^{\dagger}$	0.74	2.01	0.79	1.95	0.79
Conduct problems	4.10**	0.64*	$2.59^{\dagger}$	0.76	$2.53^{\dagger}$	0.76
Pubertal development[c]	1.52*	0.87	1.54*	0.87	1.51*	0.87
Neighborhood context						
Neighborhood poverty	0.79	1.08	$0.24^{\dagger}$	1.70	0.25	1.67
Neighborhood Context × Childhood Risk						
Neighborhood Poverty × Low SES			2.63*	0.60**	2.63*	0.60**
Neighborhood Poverty × Conduct Problems			5.21*	0.56	$4.71^{\dagger}$	0.57
Peer characteristics						
Deviant peers					$1.51^{\dagger}$	
Older partner					$2.16^{\dagger}$	

Note. OR = odds ratio; SES = socioeconomic status.
[a] Dummy markers of age were included, but associated intercepts are not shown. [b] One standard deviation below mean. [c] One standard deviation above mean.
$^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$.

Discussion

The goal of this study was to evaluate, within a developmental ecological framework, the link between the neighborhood poverty and the timing of first sexual activity in a sample of young adolescents. Results showed that ecological risk, namely, neighborhood disadvantage, was associated with early sexual initiation, but only for a subgroup of vulnerable adolescents. In contrast, the majority of young adolescents living in disadvantaged neighborhoods were not more likely to transition early. Importantly, neighborhood effects depended upon adolescents' developmental stage, gender, and background characteristics. Thus, the moderation hypotheses were generally supported. Consistent with the proposed peer mediation model, results also suggested that peer characteristics partly explained the increased risk for vulnerable young adolescents living in disadvantaged neighborhoods. In short, the confluence of developmental and ecological vulnerability was found to increase the risk of engaging in early sexual activity and peer characteristics appeared to partly explain this risk.

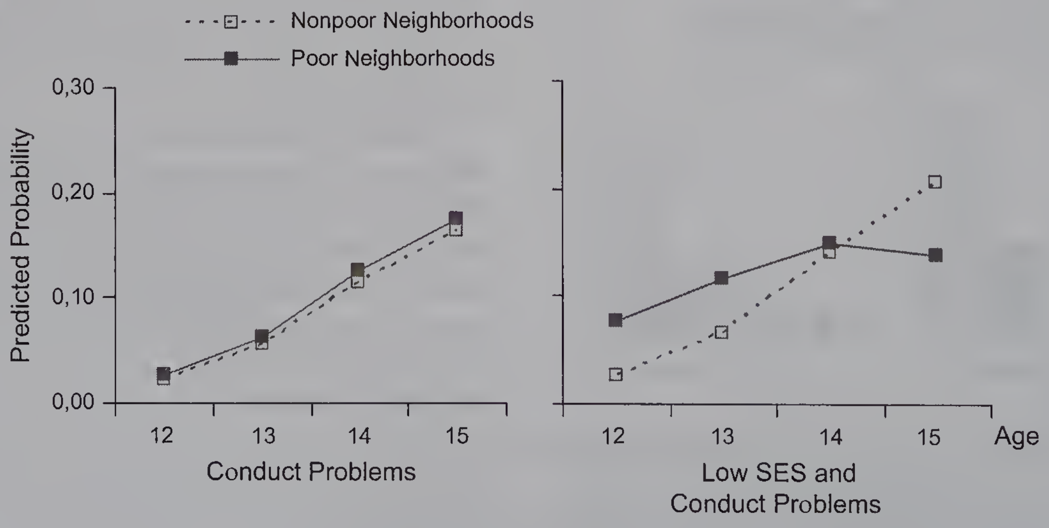


Figure 3. Adolescent males' predicted probabilities of transition to first sexual activity among those with conduct problems and among those with conduct problems and coming from a low-socioeconomic-status (SES) family, in poor and nonpoor neighborhoods, as a function of age.

For adolescent females, living in a poor neighborhood was associated with early initiation, but only for those who had a history of conduct problems. This effect was particularly strong in early adolescence, a time when sexual initiation is off-norm, and gradually faded as adolescent females progressed toward midadolescence, when it becomes more normative. Results suggested that adolescent females with conduct problems were especially at risk if they lived in poor neighborhoods possibly because in these contexts, they were more likely to affiliate with deviant, older males. In contrast, adolescent males living in poor neighborhoods were not, on average, more likely to experience early sexual initiation. However, results showed that if they had a history of conduct problems and were from a low-SES family, younger adolescent males living in poor neighborhoods were more vulnerable, as compared with adolescent males in more affluent neighborhoods with a similar profile. Among males, peer characteristics did not account for these interaction effects. Other mechanisms need to be considered in future research, such as the quality of neighborhood institutional resources or of parent–child relationships (Leventhal & Brooks-Gunn, 2000).

Consistent with Whitbeck et al.'s (1999) developmental model, many risk factors, including neighborhood disadvantage, had a stronger accelerating effect in early adolescence as compared with middle adolescence. The moderating impact of age is possibly due to the changing meaning of sexual activity across different developmental stages. Risk factors representing a departure from the norm, such as affiliation with deviant peers, predicted sexual initiation in younger adolescents, when this transition is clearly off-norms, but less so in older adolescents, when the transition becomes more normative. Interestingly, the moderating impact of age was generally stronger and more consistent among adolescent females than males, possibly because early sexual initiation represents a greater breach of social norms for them (Benda & DiBlasio, 1994). In addition to the changing meaning of sexual initiation as a function of age, other developmental processes might explain younger adolescents' increased sensitivity to neighborhood effects. For instance, as compared with older peers, younger adolescents are thought to be more susceptible to peer pressure and perceived social norms (Brooks-Gunn & Furstenberg, 1989).

In this study, neighborhood disadvantage had a direct, independent accelerating effect in adolescent females but not in males. This result is not wholly consistent with previous findings in the neighborhood literature. A study conducted with a sample of Black adolescents found an independent effect of neighborhood disadvantage only in males (Ramirez-Valles et al., 2002). Another study conducted in a racially diverse sample found comparable neighborhood effects in both genders (Browning et al., 2005). In contrast, the present study, using a virtually all-White sample, found significant direct neighborhood effects only in adolescent females. Gender differences in neighborhood effects appear to depend on adolescents' racial and cultural background perhaps because gender-specific norms regarding the optimal timing for entry into sexual activity vary considerably across groups (East, 1998). Because cultural norms prohibiting early sexual activity tend to be especially emphasized for White adolescent females, early sexual activity represents a greater breach of norms for them compared with other racial groups (Cavanagh, 2004). This could explain why, in this sample, peer group characteristics marking an evident departure from the norm, such as peer deviance, were strongly associated with early sexual activity for adolescent females but not for males. This interpretation is also supported by previous studies showing that peer norms are powerfully associated with White adolescent females' sexual behavior, an association that appears specific to adolescent females from this racial group (Billy & Udry, 1985). Clearly, additional research is needed to uncover the mechanisms underlying gender differences in neighborhood effects within different racial and ethnic groups.

Among adolescent females, the impact of preexisting conduct problems was exacerbated in disadvantaged neighborhoods, apparently because these adolescent females were more at risk for affiliating with older, deviant males if they lived in a poor neighborhood. At least two explanations might account for this effect. First, adolescent females with such a background might seek the company of peers with a similar behavioral profile. In disadvantaged neighborhoods, this selection process is potentially facilitated given the higher prevalence of problem behaviors (Sampson et al., 2002). By seeking the company of similar peers, these adolescent females might often affiliate with older, deviant males because as a group, older adolescent males exhibit the highest level of delinquent behaviors (Hirschi & Gottfredson, 1983). Reciprocation could reinforce this process, with older deviant males likely to find younger adolescent females attractive (Caspi et al., 1993). In this social context, peer norms concerning the desirability of sexual activity, as well as developmental imbalances, are potentially responsible for the increased likelihood of sexual initiation. Second, family processes might also play an important role. For instance, adolescents with conduct problems are especially impervious to parents' supervision efforts and inconsistent supervision is more strongly associated with early sexual initiation in disadvantaged neighborhoods (Browning et al., 2005). Thus, friendship preferences and family dynamics, in combination with an increased accessibility of deviant peers, might explain why in disadvantaged neighborhoods young adolescent females with a history of conduct problems are especially likely to engage in early sexual activity.

Finally, neighborhood poverty also amplified the effect of family disadvantage on early initiation. Adolescent males with a history of conduct problems and coming from a low-SES family were more likely to become sexually initiated in early adolescence, but only if they also lived in a poor neighborhood, and among adolescent females, those living in a poor neighborhood and originating from a residentially unstable family were marginally more at risk. Different mechanisms, in addition to the supervision pathway discussed above, may explain why the negative impact of family disadvantage appears reinforced in poor neighborhoods. First, adolescents from a disadvantaged family background may be more likely to develop a sense of hopelessness when their families' difficulties are echoed in their larger environment (Wilson, 1987). In turn, "nothing to lose" attitudes are associated with adolescent involvement in risk behaviors (Harris, Duncan, & Boisjoly, 2002). Also, residential instability may further reduce access to the already scarce social and institutional resources in disadvantaged neighborhoods. Notably, residential instability has a negative impact on the quality of adolescents' friendship networks. Indeed, adolescents who have recently moved to a new neighborhood have been found to be more likely to affiliate with deviant friends, which in turn increases the risk for early sexual activity (South et al., 2005).

The conclusions of this study should be considered with some important limitations in mind. First, even if special care has been devoted to circumvent possible bias due to attrition, limits to generalizability may remain. In particular, as noted above, the results might apply only to the cultural group studied. Second, even if important family and individual confounding factors were taken into account, the possibility of selection bias can never be fully ruled out. Indeed, there is always a chance that an omitted variable associated with both neighborhood poverty and sexual initiation is responsible for neighborhood effects. Third, as is often the case in large-scale studies investigating many aspects of development, some of the concepts investigated would have benefited from more thorough measurement strategies. For instance, in addition to retrospective information about the age of first sexual partners, prospective information about the age of boyfriends and girlfriends would have strengthened our results and refined our interpretations. In the same vein, additional details about the circumstances of first intercourse, such as the degree of consent, would have allowed us to make additional distinctions. Fourth, the survey questions did not allow distinctions between sexual initiations with same- and other-gender partners. For this reason, the explanatory model proposed, which is centered on other-gender sexual activity, misrepresents the experiences of adolescents engaging in same-gender sexual activities. In addition, even if the sample in this study is large, the main conclusions are based on relatively small numbers of individuals. Indeed, because the NLSCY is a normative sample, early intercourse was relatively rare. Replications of findings in other studies oversampling at-risk adolescents are warranted. Also, although results from prediction models are robust, point estimates of the proportion of sexually experienced adolescents at different ages can vary appreciably depending on how missing data and data inconsistencies are treated and for this reason should be viewed with caution (Upchurch, Lillard, Aneshensel, & Fang Li, 2002). Finally, because adolescent males are more likely than adolescent females to provide inconsistent reports of age at sexual initiation (Upchurch et al., 2002), some of the differing results observed for adolescent males and adolescent females might reflect this tendency. Notably, it might contribute to the lack of significant predictors of adolescent males' sexual behavior.

Another important limitation of the study is the uncertainty regarding the temporal order of the peer explanatory variables. It was suggested that adolescent females with conduct problems were more likely to affiliate with older, deviant males during early adolescence and that this tendency in turn placed them at risk for early sexual activity. However, because the peer processes were measured concurrently, it is equally possible that adolescent females were introduced into deviant peer groups through early sexual debut. Thus, because of the correlational nature of this study and because of the uncertain temporal order of the explanatory processes, causation cannot be directly inferred. More detailed interview data would be necessary to grasp the temporal ordering of the events leading to early sexual activity in various contexts.

This study used a developmentally oriented framework for examining the link between the neighborhood disadvantage and the timing of sexual initiation. In line with previous studies showing that the association between neighborhood disadvantage and early sexual activity depends on family characteristics

(Browning et al., 2005), this study further demonstrated that this association also depends on individual characteristics such as age and behavioral traits. By identifying young adolescents who are particularly at risk within disadvantaged communities, this study provides valuable insights for future intervention efforts. The results suggest that to maximize effectiveness, prevention programs need to take the larger social context into account and make special efforts to enroll vulnerable young adolescents. Also, the results suggest that additional research is needed that examine the role of neighborhood peers, a potentially important explanatory factor for neighborhood effects on adolescent sexual behavior. If future research confirms a causal role, prevention programs aimed at delaying first sexual activity could benefit from integrating strategies for reducing affiliation with deviant neighborhood peers.

References

Albert, B., Brown, S., & Flanigan, C. (Eds.). (2003). *14 and younger: The sexual behavior of young adolescents (Summary).* Washington, DC: National Campaign to Prevent Teen Pregnancy.

Allison, P. D. (2001). *Missing data.* Thousand Oaks, CA: Sage.

Baumer, E. P., & South, S. J. (2001). Community effects on youth sexual activity. *Journal of Marriage and Family, 63,* 540 – 554.

Benda, B. B., & DiBlasio, F. A. (1994). An integration of theory: Adolescent sexual contacts. *Journal of Youth and Adolescence, 23,* 403 – 420.

Billy, J. O. G., & Udry, J. R. (1985). The influence of male and female best friends on adolescent sexual behavior. *Adolescence, 20,* 21 – 32.

Brody, G. H., Ge, X., Conger, R. D., Gibbons, F. X., Murry, V. M., Gerrard, M., et al. (2001). The influence of neighborhood disadvantage, collective socialization, and parenting on African American children's affiliation with deviant peers. *Child Development, 72,* 1231 – 1246.

Bronfenbrenner, U. (1979). *The ecology of human development.* Cambridge, MA: Harvard University Press.

Brooks-Gunn, J., & Furstenberg, F. F. (1989). Adolescent sexual behavior. *American Psychologist, 44,* 249 – 257.

Browning, C. R., Leventhal, T., & Brooks-Gunn, J. (2004). Neighborhood context and racial differences in early adolescent sexual activity. *Demography, 41,* 697 – 720.

Browning, C. R., Leventhal, T., & Brooks-Gunn, J. (2005). Sexual initiation in early adolescence: The nexus of parental and community control. *American Sociologial Review, 70,* 758 – 778.

Caspi, A., Lynam, D. R., Moffitt, T. E., & Silva, P. A. (1993). Unraveling girls' delinquency: Biological, dispositional, and contextual contributions to adolescent misbehavior. *Developmental Psychology, 29,* 19 – 30.

Cavanagh, S. E. (2004). The sexual debut of girls in early adolescence: The intersection of race, pubertal timing, and friendship group characteristics. *Journal of Research on Adolescence, 14,* 285 – 312.

Darroch, J. E., Frost, J. J., Singh, S., & Study Team. (2001). *Teenage sexual and reproductive behavior in developed countries: Can more progress be made?* Occasional Report, The Alan Guttmacher Institute, New York.

Dickson, N., Paul, C., Herbison, P., & Silva, P. (1998). First sexual intercourse: Age, coercion, and later regrets reported by a birth cohort. *British Medical Journal, 316,* 29 – 33.

Dorn, L. D., Susman, E. J., & Ponirakis, A. (2003). Pubertal timing and adolescent adjustment and behavior: Conclusions vary by rater. *Journal of Youth and Adolescence, 32,* 157 – 167.

East, P. L. (1998). Racial and ethnic differences in girls' sexual, marital, and birth expectations. *Journal of Marriage and the Family, 60,* 150 – 162.

French, D. C., & Dishion, T. J. (2003). Predictors of early initiation of sexual intercourse among high-risk adolescents. *Journal of Early Adolescence, 23,* 295 – 315.

Garriguet, D. (2005). Early sexual intercourse. *Health Reports (Statistics Canada, Catalogue 82-003), 16,* 9 – 18.

Graber, J. A., Brooks-Gunn, J., & Galen, B. R. (1998). Betwixt and between: Sexuality in the context of adolescent transitions. In R. Jessor (Ed.), *New perspectives on adolescent risk behavior* (pp. 270 – 316). Cambridge, UK: Cambridge University Press.

Halpern, C. T., Kaestle, C. E., & Hallfors, D. D. (2007). Perceived physical maturity, age of romantic partner, and adolescent risk behavior. *Prevention Science, 8,* 1 – 10.

Hansen, L., Mann, J., McMahon, S., & Wong, T. (2004). Sexual health. *BMC Women's Health, 4*(Suppl. 1), S24 – S32.

Harding, D. J. (2007). Cultural context, sexual behavior, and romantic relationships in disadvantaged neighborhoods. *American Sociologial Review, 72,* 341 – 364.

Harris, K. M., Duncan, G. J., & Boisjoly, J. (2002). Evaluating the role of "nothing to lose" attitudes on risky behavior in adolescence. *Social Forces, 80*(3), 1005 – 1039.

Hirschi, T., & Gottfredson, M. (1983). Age and the explanation of crime. *American Journal of Sociology, 89,* 552 – 584.

Ingoldsby, E. M., Shaw, D. S., Winslow, E., Schonberg, M., Gilliom, M., & Criss, M. M. (2006). Neighborhood disadvantage, parent-child conflict, neighborhood peer relationships, and early antisocial behavior problem trajectories. *Journal of Abnormal Child Psychology, 34,* 293 – 319.

Kaestle, C. E., Halpern, C. T., Miller, W. C., & Ford, C. A. (2005). Young age at first intercourse and sexually transmitted infections in adolescents and young adults. *American Journal of Epidemiology, 161,* 774 – 780.

Leventhal, T., & Brooks-Gunn, J. (2000). The neighborhoods they live in: The effects of neighborhood residence in child and adolescent outcomes. *Psychological Bulletin, 126,* 309 – 337.

Manlove, J., Ryan, S., & Franzetta, K. (2007). Risk and protective factors associated with the transition to a first sexual relationship with an older partner. *Journal of Adolescent Health, 40*, 135–143.

Marin, B. V., Coyle, K. K., Gomez, C. A., Carvajal, S. C., & Kirby, D. B. (2000). Older boyfriends and girlfriends increase risk of sexual initiation in young adolescents. *Journal of Adolescent Health, 27*, 409–418.

Maticka-Tyndale, E., McKay, A., & Barrett, M. (2001). *Teenage sexual and reproductive behavior in developed countries: Country report for Canada*. New York: The Alan Guttmacher Institute.

Obeidallah, D., Brennan, R. T., Brooks-Gunn, J., & Earls, F. (2004). Links between pubertal timing and neighborhood contexts: Implications for girls' violent behavior. *Journal of the Academy of Child and Adolescent Psychiatry, 43*, 1460–1468.

Pepler, D. J., & Sedighdeilami, F. (1998). *Aggressive girls in Canada*. Ottawa, Ontario, Canada: Human Resources Development Canada, Applied Research Branch.

Puderer, H. (2001). Introducing the dissemination area for the 2001 census: An update. In *Geography Working Paper Series* (Vol. 2000–2004). Ottawa, Canada: Statistics Canada.

Ramirez-Valles, J., Zimmerman, M. A., & Juarez, L. (2002). Gender differences of neighborhood and social control process: A study of the timing of first intercourse among low-achieving, urban, African American youth. *Youth and Society, 33*, 418–441.

Sampson, R. J., Morenoff, J. D., & Gannon-Rowley, T. (2002). Assessing "neighborhood effects": Social processes and new directions in research. *Annual Review of Sociology, 28*, 442–478.

Shapka, J. D., Dahinten, V. S., & Arım, R. G. (2006). *Adolescent early sexual activity: An examination of risk and protective factors*. Paper presented at the European Association for Research on Adolescence, Antalya, Turkey.

Simons, R. L., Johnson, C., Beaman, J., Conger, R. D., & Whitbeck, L. B. (1996). Parents and peer group as mediators of the effect of community structure on adolescent problem behavior. *American Journal of Community Psychology, 24*, 145–171.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford, UK: Oxford University Press.

South, S. J., Haynie, D. L., & Bose, S. (2005). Residential mobility and the onset of adolescent sexual activity. *Journal of Marriage and Family, 67*, 499–514.

Statistics Canada. (2005). *National Longitudinal Survey of Children and Youth microdata user guide—Cycle 5*. Ottawa, Canada.

Udry, J. R., & Billy, J. O. G. (1987). Initiation of coitus in early adolescence. *American Sociological Review, 52*, 841–855.

Upchurch, D. M., Aneshensel, C. S., Sucoff, C. A., & Levy-Storms, L. (1999). Neighborhood and family contexts of adolescent sexual activity. *Journal of Marriage and the Family, 61*, 920.

Upchurch, D. M., Lillard, L. A., Aneshensel, C. S., & Fang Li, N. (2002). Inconsistencies in reporting the occurrence and timing of first intercourse among adolescents. *Journal of Sex Research*, 197–206.

Whitbeck, L. B., Yoder, K. A., Hoyt, D. R., & Conger, R. D. (1999). Early adolescent sexual activity: A developmental study. *Journal of Marriage and Family, 61*, 934–946.

Willms, J. D., & Shields, M. (1996). *A measure of socioeconomic status for the National Longitudinal Study of Children*. Fredericton: Canadian Research Institute for Social Policy.

Wilson, W. J. (1987). *The truly disadvantaged: The inner city, the underclass, and public policy*. Chicago: University of Chicago Press.

Appendix

Table A1

Design Characteristics: Timing of Measurement and Informant

	Timing of measurement		
	Type	Age (years)	Informant
Socioeconomic status	Antecedent	10–11	Parent
Intact family status	Antecedent	10–11	Parent
Family residential instability	Antecedent	10–11	Parent
Conduct problems	Antecedent	10–11	Self
Pubertal development	Antecedent	10–11	Self
Neighborhood poverty	Concurrent	12–13 or 14–15	Census
Deviant peers	Concurrent	12–13 or 14–15	Parent
Older partner	Retrospective	16–17 or 18–19	Self
Age at first consensual intercourse	Retrospective	16–17 or 18–19	Self

Child Development, September/October 2008, Volume 79, Number 5, Pages 1477–1497

Searching and Planning: Young Children's Reasoning About Past and Future Event Sequences

Kerry L. McColgan and Teresa McCormack
Queen's University Belfast

Six experiments examined children's ability to make inferences using temporal order information. Children completed versions of a task involving a toy zoo; one version required reasoning about past events (search task) and the other required reasoning about future events (planning task). Children younger than 5 years failed both the search and the planning tasks, whereas 5-year-olds passed both (Experiments 1 and 2). However, when the number of events in the sequence was reduced (Experiment 3), 4-year-olds were successful on the search task but not the planning task. Planning difficulties persisted even when relevant cues were provided (Experiments 4 and 5). Experiment 6 showed that improved performance on the search task found in Experiment 3 was not due to the removal of response ambiguity.

The development of flexible and imaginative thinking about nonpresent events has been the subject of recent attention. This cognitive ability has been described in many ways: for example, as episodic past and future thinking (Atance & O'Neill, 2001a), temporal decentering (Zelazo & Somerville, 2001), or temporal perspective taking (McCormack & Hoerl, 1999), but most notably as mental time travel (MTT; Suddendorf & Busby, 2003). Although there are differences between these descriptions, all these theorists emphasize the fact that there is a flexible type of thinking about events in the past and the future that is firmly established relatively late in development (around the age of 4–5). Despite the recent interest in the development of MTT, it remains the case that there have been relatively few attempts to characterize important, and perhaps associated, changes in the nature of temporal concepts and reasoning in the early years (for notable exceptions, see Nelson, 1996; Friedman, 2003; Weist, 1989).

According to our everyday conception of time, time is populated by a series of chronologically ordered events stretching from the present in two directions, into the past and into the future. Povinelli (2001; Povinelli, Landry, Theall, Clark, & Castille, 1999) has argued that mature temporal cognition requires not only the ability to chronologically order events in time but also an appreciation of the causal flow of events. Our everyday thinking about events in time is based on the awareness that chronological order is a result of "an inherent causal arrow of time that binds each state of the world to each succeeding one" (Povinelli et al., 1999, p. 1427). The temporal order in which events happen often affects the overall outcome of a sequence of events, due to the unidirectional nature of the causal arrow of time (earlier events will have an influence on later ones but not vice versa) and the causal relations between these events (later events in a sequence may change or remove the effects of events that occurred earlier in the sequence).

Flexible and insightful temporal cognition involves grasping the actual practical significance of this chronological organization, a significance that derives from its causal structure. This type of reasoning has sometimes been referred to as temporal–causal reasoning (McCormack & Hoerl, 2005; see also Welch-Ross, 2001). It is clear that from an early age, children are able to represent and remember temporal order information: For example, children younger than 2 years can reproduce novel action sequences in the correct order (Bauer & Mandler, 1989), and by 3 or 4 years, they can correctly order representations of sequences of familiar events, such as the main events of a typical day or a visit to the supermarket (Benson, 1997; Fivush & Mandler, 1985; Friedman, 1977). However, the extent to which preschool children can incorporate information about temporal order into their reasoning is less clear. Povinelli et al.'s (1999) argument is that the ability to represent and

---

Kerry L. McColgan is now at the University of Warwick, United Kingdom.

This work was supported by a PhD studentship awarded to the first author by the Department of Education and Learning, Northern Ireland, and by a grant from the Arts and Humanities Research Council awarded to the second author.

Correspondence concerning this article should be addressed to Kerry L. McColgan, School of Psychology, University of Warwick, Coventry, CV4 7AL, United Kingdom. Electronic mail may be sent to k.l.t.mccolgan@warwick.ac.uk.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7905-0017

remember such order information may be developmentally more primitive than the ability to use it to support inferences that require a consideration of event order. For example, can young children use information about the order in which a sequence of novel events have happened to make inferences about how the world must be now?

We have described this type of reasoning in question as temporal–causal reasoning, with the term *causal* used because such reasoning involves a consideration of the outcome or consequences of the order in which events have occurred or will occur. However, most of the developmental research that has been conducted on causal reasoning has focused on what might be termed causal powers reasoning, for example, whether a block has the causal power to make a machine light up and play a tune (Gopnik, Sobel, Schulz, & Glymour, 2001; for a review, see Corrigan & Denton, 1996). Temporal–causal reasoning may build on existing knowledge about causal powers, but temporal–causal reasoning tasks place an additional demand on children's inferential abilities over and above causal powers reasoning. Specifically, children need to not only consider the causal powers of events but also reason about the role played by event order in determining outcome.

Povinelli et al. (1999) and McCormack and Hoerl (2007) have examined this type of reasoning using simple object location tasks. In both their studies, an object was placed out of sight of the child, first in one location and then in another. At test, children were unexpectedly asked to locate the object but could only do so correctly by retrospectively considering the order of the two hiding events they had not previously observed. Children aged 5 years or younger had consistent difficulties with doing so. McCormack and Hoerl (2005) also examined preschool children's ability to reason about the causal significance of the order in which two events occurred. In their study, children were required to judge which of two objects could be found in the window of a novel mechanical box. Answering correctly required a retrospective consideration of the order in which two button presses involving the box had occurred. Again, children younger than 5 years performed at chance levels on this task. These recent studies suggest that basic temporal–causal reasoning abilities are still developing between 4 and 5 years and provide support for Povinelli et al.'s suggestion that children can represent order information before being able to exploit it in their reasoning. However, this research has focused solely on reasoning involving short sequences of events that occurred in the past and has not examined reasoning involving event sequences located in the future.

In the experiments reported, we designed tasks to measure temporal–causal reasoning about event sequences for both the past and the future. The tasks involved mentally revisiting past events or imagining future events and considering the causal consequences of the order in which such events had occurred (past version of the task) or were to occur (future version). As described and explained below, a search task was used as a measure of temporal–causal reasoning concerning the past and a simple planning task as a measure of reasoning concerning the future. Based on a suggestion by Goswami (1998), both versions involved a visit of a doll to a toy zoo that carried a camera.

Reasoning About Past Event Sequences

Although relatively few studies have explicitly examined the development of temporal–causal reasoning, some previously established tasks in the literature can be easily interpreted as placing demands on this type of reasoning even though they have been used to investigate other cognitive abilities. For example, in some previous studies of children's search behavior, searching has taken place after children have visited a series of locations (e.g., Anooshian, Hartman, & Scharf, 1982; Haake, Somerville, & Wellman, 1980; Wellman, Somerville, & Haake, 1979). Participants' visits to each location and the activities carried out at locations can be viewed as a sequence of events that has unfolded over time. Typically, the tasks have involved locating an object that has previously been used during an activity at a certain point in the sequence, with children being unaware in advance that they would be required to locate the object. Children were thus required to think retrospectively about the order in which events in the sequence had occurred and use such a consideration to infer a state of the world—specifically, the possible location of the object. Clearly, such search tasks load on a variety of cognitive abilities, in particular memory and spatial abilities. However, some of these tasks also require temporal–causal reasoning, in that accurate performance depends on a consideration of the causal significance of the order in which events have occurred. For example, Wellman et al.'s (1979) task involved children who, after visiting a series of locations, had to search for a missing camera that had been used at a specific location in the sequence. Although the youngest children were often able to correctly return to the location where the picture had been taken, only the older children were consistently able to infer the only other logically possible locations of the object. This suggests that younger children may just have looked in the place the camera had last been used based on its association with that location. Although other studies have attempted to rule out such an associative explanation (Anooshian et al., 1982; Haake et al., 1980), their findings do not straightforwardly permit the conclusion that young children realize that only certain locations are logically possible, given the sequence of events that has unfolded.

Our task resembled such tasks in that it involved a single trial, thus eliminating the possibility of perseveration at certain search locations across trials (see Suddendorf, 2003). Furthermore, as in these previous search tasks, children were unaware in advance that they would need to search for an object, ruling out the possibility that responses could be based on simple object tracking rather than retrospective reasoning, However, our search task differed from these previous tasks in a number of ways. First, a simple associative response (in the case of our task, choosing the location where the camera was last used) was not possible. Second, children were given only a single opportunity to choose a location so that responses could be easily interpreted and classified. Third, most previous search studies have been conducted in large-scale relatively complex spatial environments; in our task, the spatial layout was deliberately extremely simple. Last, given that Wellman et al. (1979) found that younger children were not always able to recall the last location in which an object (also a camera) was used, we provided children with a memory cue in the form of a Polaroid photograph that had been taken with the camera. Furthermore, the fact that temporal order was mirrored in the spatial layout available to the child at test meant that children did not have to conduct an extensive memory search to reconstruct the sequence of events, reducing the possibility of failure due to memory rather than reasoning problems. It was hoped that these various modifications would ensure that children who failed the task would do so because of poor temporal – causal reasoning abilities rather than because of deficits in other cognitive skills such as memory or spatial cognition.

Reasoning About Future Event Sequences

Children's ability to consider the order of event sequences in the future has previously been studied using a wide variety of planning tasks (Bauer, Schwade, Wewerka, & Delaney, 1999; Benson, 1997; Hudson & Fivush, 1991; Hudson, Sosa, & Shapiro, 1997; Wellman, Fabricius, & Sophian, 1985; see Scholnick, Friedman, & Wallner-Allen, 1997, for a review). Atance (Atance & O'Neill, 2001b) has developed planning tasks for use with young preschoolers, and her findings have suggested that at least some 3-year-olds can make simple plans. However, the tasks that she has used are typically one-step planning tasks that do not actually require a consideration of temporal order of a future sequence of events. In contrast, traditional formal planning tasks such as the Tower of London/Hanoi task require a consideration of the order in which actions in a sequence must be executed (Scholnick et al., 1997). Some studies have used versions of such traditional planning tasks to examine development (e.g., Klahr & Robinson, 1981; Welsh, 1991), although these tasks have not always proved straightforward to administer to younger children (Luciana & Nelson, 1998). In general, their findings suggest marked improvements between 3 and 5 years (e.g., Carlson, Moses, & Claxton, 2004; Welsh, 1991), in line with findings from planning tasks set in more naturalistic contexts (Hudson et al., 1997). The planning task used in the present experiments differed from many previous planning tasks (e.g., those employing the Tower of London/Hanoi) in that children were deliberately given only a single trial, without any initial demonstrations by the experimenter or the opportunity to improve their performance over a number of trials. Thus, performance could not be based on acquired strategies or routines. Furthermore, in contrast to some planning tasks (see Hudson et al., 1997), the event sequence was entirely novel so that children could not use their knowledge of either familiar event sequences or a sequence they had recently observed to answer correctly.

Comparing Past and Future Reasoning

A common feature of the theoretical accounts mentioned at the start of this article is that they are intended to encompass cognition about both the past and the future. For example, on Suddendorf's account, MTT to the past and MTT to the future are skills that would be expected to go hand in hand because they involve a common core cognitive ability (see also Atance & O'Neill, 2005). Despite this, there is little research that has directly compared whether the development of such skills emerges simultaneously. Neuropsychological research has begun to examine the extent to which reasoning about the past and imagining the future involve common brain areas (e.g., Hassabis, Kumaran, Vann, & Maguire, 2007; Szpunar, Watson, & McDermott, 2007). Moreover, recent comparative research has explored whether certain animal species are capable of both thought about the past and the future (e.g., Clayton, Bussey, & Dickinson, 2003; Raby, Alexis, Dickinson, & Clayton, 2007). However, relatively little research has attempted to directly compare the development of temporal cognition concerning the past with the development of that concerning the future, although this issue is beginning to attract attention (Busby & Suddendorf, 2005; Hudson, 2006).

Busby and Suddendorf (2005) attempted to address this issue by using a verbal task to elicit both past and future talk from young children. They found that 4- and 5-year-olds could answer questions about both the past and the future equally well and found no age-related differences. They take their findings as a demonstration that the ability to engage in MTT to the future and to the past emerge in tandem. In contrast, Hudson (2002, 2006) has argued that preschool children find it more difficult to engage in conversation about the future than about the past. Hudson has also argued that the ability to remember event sequences may precede the ability to use such memories to flexibly plan for the future (Hudson et al., 1997; see also Benson, 1997). Other related research conducted by Friedman (for a summary, see Friedman, 2003, 2005) has examined young children's abilities to judge the relative distance from the present of both past and future events. Looking across the findings of his studies, it would appear that the ability to make judgments of this sort for future events appears to develop somewhat later than for past events. Indeed Friedman (2005) has argued that, at least in young children, qualitatively different types of processes may underpin such judgments for the past and the future.

The series of experiments described below examined the development of temporal–causal reasoning for both past and future sequences of events, using our simple search and planning tasks. If these tasks measure a common ability, then we might expect to see similar developmental patterns for both the tasks. Experiments 1 and 2 examined how performance changed between 3 and 5 years on the two tasks. In Experiments 3–5, we examined whether performance could be improved by using a shorter sequence of events, and Experiment 6 explored possible explanations of the effect on performance of reducing the number of events in the sequence in the search task.

Experiment 1

Method

Participants

One hundred and eighty children took part in the task; 60 in each of the three age groups tested, with half in each age group taking part in the search task and half completing the planning task. The groups comprise sixty 3-year-olds (search: M = 44 months, range = 40–47 months; planning: M = 45 months, range = 41–47 months), sixty 4-year-olds (search: M = 55 months, range = 51–59 months; planning: M = 55 months, range = 49–59 months), and sixty 5-year-olds (search: M = 68 months, range = 60–71 months; planning: M = 68 months, range = 60–71 months). There were 85 boys and 95 girls. All children were of Caucasian descent and were mainly from working- and middle-class backgrounds. Participants were drawn from nurseries and schools local to the first author. For both versions of the task, children were tested individually in a quiet area away from the other children in the group or class. The tasks took approximately 10 min each, and all children received a set of stickers for taking part.

Materials

The same novel zoo apparatus was used for both tasks (see Figure 1a). This comprised a large rectangle of green felt on which was sewn a gray felt semicircular path. Five cages, with a different miniature plastic animal in each, were placed at regular intervals along the curve of the path. These animals were a giraffe, zebra, kangaroo, elephant, and a tiger. The cages were of different dimensions (between 17 × 11 cm and 23 × 16 cm) to accommodate the different sizes of the animals. Four boxes (referred to as lockers), each measuring 21 × 24 cm, were placed behind the cages of the animals at the first, second, fourth, and fifth locations visited along the path. These lockers were hinged at the top, which enabled them to be opened up. During the task, the lockers were placed in such a way that they opened away from the experimenter, ensuring that participants were unable to see into them. A Polaroid camera, measuring 18 × 5 cm, was used in both tasks, and a soft female toy doll (Molly) measuring 38 cm high was the actor in the two tasks. The procedure of the search task involved the doll's camera being "lost" at either Location 4 or 5 along the path. In order to achieve this without the child's knowledge, a second duplicate camera was used. This meant that one camera could be hidden in an appropriate locker from the outset of the task, whereas the other was carried by the doll during the task.

A green felt rucksack measuring 25 × 22 cm was designed to hold the camera that was used during the task, as well as additional items. The rucksack fitted onto the doll's back. Two identical versions of this were required. One of these bags (the original bag)

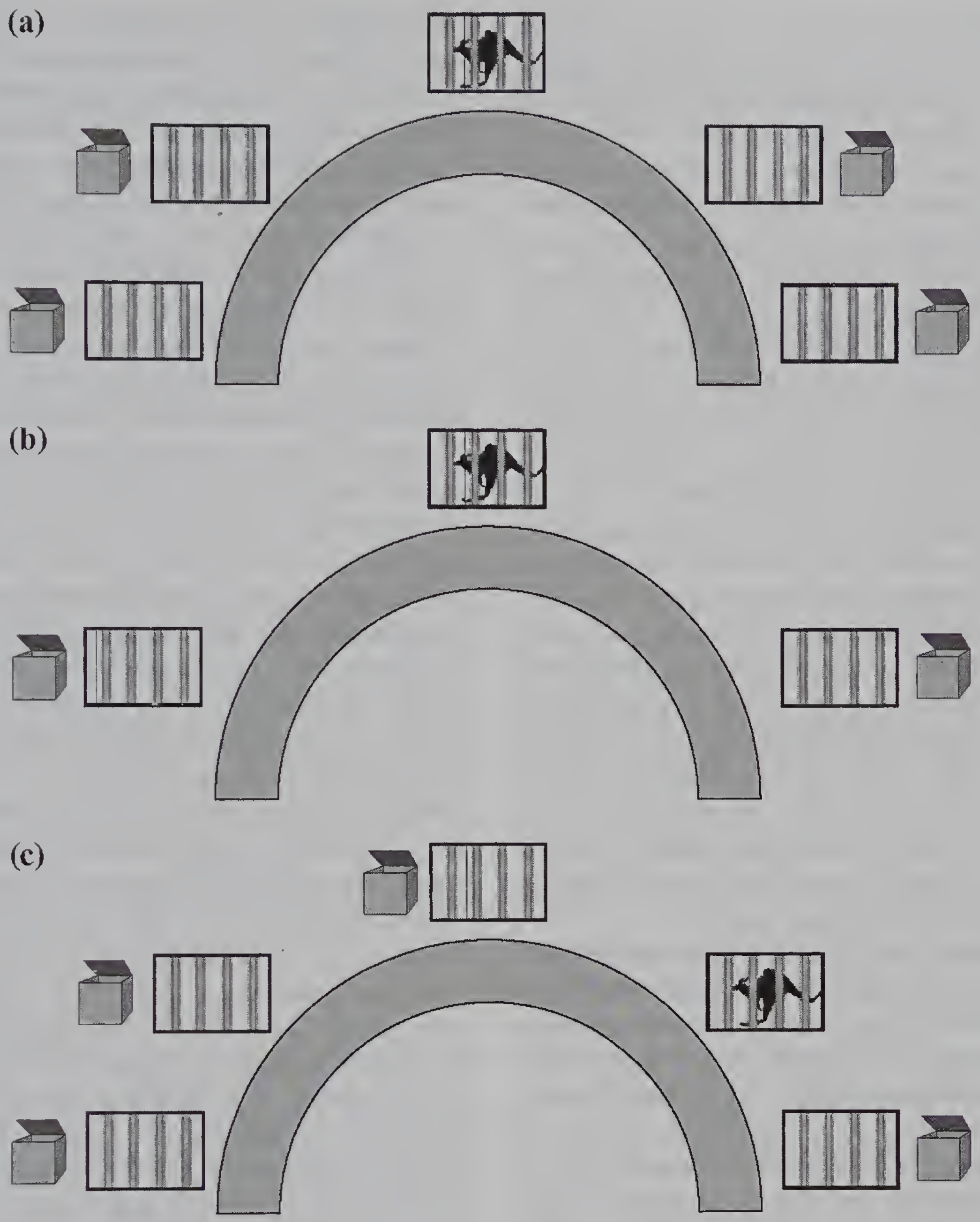


Figure 1. Schematic drawing of the layout used in (a) Experiments 1 and 2, (b) Experiments 3–5, and (c) Experiment 6.
Note. For ease of visualization of the array, boxes are shown in this figure positioned as if the child could see inside each of them when opened. However, in all studies, the boxes were rotated 180°, so that the child could not see inside any of them when opened by the experimenter.

contained all the items listed below and the camera. The other bag contained all the items listed below but no camera (the duplicate bag). The additional items were (one for each of the bags): a toy dog, a notebook, and a small bottle of juice. The rucksacks were constructed such that when closed, the contents of the bag could not be discerned; thus, it was impossible to tell whether or not the bag did or did not contain the camera simply from viewing its exterior. A doll's coat was used as part of the distracter task. The order in which the animals were placed along the zoo's path was varied in order to control for the possibility that a child's response may have been influenced by the location of their favorite animal. There was one restriction to these variations, which was that the kangaroo's cage was always placed at Location 3, which was the middle location in this study.

Procedure

At the beginning of the task, children were seated in front of the zoo, directly facing the middle cage, and equidistant from all the other locations.

Search task. Prior to the start of the task, before children were present, the second camera was placed by the experimenter in the box at either Location 4 or 5 along the path (this was counterbalanced). It remained there throughout the experiment, although children could not see it throughout the procedure. The duplicate bag was also hidden in a suitable place somewhere out of view of children, where it could be quickly retrieved by the experimenter. Additionally, in order to provide a brief distracter task, the doll's coat was positioned somewhere just outside the testing room, so that in order to retrieve it, children would have to leave the room. This gave the experimenter time to swap the original bag with the duplicate one.

When children were seated, they were introduced to Molly and told that she was at the zoo. They were given a brief description of the nature of the zoo. The experimenter told children that Molly had a bag with her and then showed them all the contents of the bag. In particular, children's attention was directed toward the Polaroid camera. They were told that this type of camera produced the photo "nearly right away." To demonstrate this, a photo was taken of Molly. To help children's understanding, the experimenter showed children the photo and explained to them that at the start, you couldn't see the picture of Molly, but after a while, the photo changed color and you could see Molly. Once children had seen the fully developed photo, they were shown the lockers beside the cages at Locations 1, 2, 4, and 5. The experimenter told children that these lockers could be opened up, and so whenever Molly visited an animal with a locker beside it, she should put her bag into it.

At this point, the experimenter explained to children that there was a path in the zoo that the doll had to go round, which the experimenter traced with their finger, going from children's right to left. They were told that Molly had to go a certain way around the zoo and that she could only go round it once and so could not go back again and revisit an animal. In order to determine if children understood these instructions, the experimenter asked them to point to the start and end of the path and to show her which way round the zoo the doll had to go. Once these details had been explained to children, the experimenter took the doll to Location 1. The bag was taken off the doll and put into the locker that was beside this cage. Children and the experimenter then talked about the animal for a short while and then the experimenter retrieved the bag from the locker, put it back on the doll's back, and then moved on to Location 2 where this sequence was repeated.

At Location 3, however, the experimenter pointed out that there was no locker beside the cage there, and therefore this meant that the bag had to stay on Molly's back. It was suggested that because Molly had her bag with her, they should take a photo of the kangaroo, which was the animal in that cage. The camera was removed from the bag and a photo was taken that was then placed in front of the kangaroo's cage, where it stayed for the remainder of the task. The experimenter then put the camera back in the bag, with children being made aware that she was doing so. The experimenter then took the doll to Locations 4 and 5 in turn, and again, at each of these locations, the bag was taken off the doll's back and placed in the respective lockers, as with Locations 1 and 2.

When the doll reached the end of the zoo and had visited all the animals, participants were given a brief distracter task. This involved them being asked to retrieve the doll's coat which, as previously mentioned, had been placed just outside of the room. While children were away from the room, the experimenter exchanged the original bag that had been used during the task (with the camera) for the duplicate bag without the camera. When children returned, the experimenter told them that the camera was missing and let them look in the bag to see for themselves. She told children that the camera must have fallen out whenever the bag was being put into one of the lockers. Children were then shown the photo of the kangaroo that had been taken during the task and told "She had it earlier because she took this nice photo of the kangaroo. Shall we help Molly find her camera?" The experimenter then asked children what locker they thought the camera might be in. They were told that it was only possible to look in one location. Children's answer was taken as the locker that they pointed to. The correct answer for this task was either Location 4 or 5 (regardless of where the duplicate camera had actually been hidden). The experimenter looked in the locker that children had indicated: If the camera was in that location, she showed them that they had chosen correctly; if it was not, she showed them the true location of the camera. Regardless of whether or not children had found the camera, they were thanked for taking part and given a sticker of their choice.

Planning task. Once children had been seated, the experimenter introduced them to the doll, Molly. In this task, they were told that, "Molly will be going to the zoo." As in the search condition, children were given a brief description of what a zoo was. The experimenter then told children, "Molly has a camera with her that she wants to bring to the zoo." The Polaroid camera was shown to them, and the experimenter explained how it worked; as in the search task, this was demonstrated by taking a photo of Molly. Children were then told that Molly did not have a bag with her and so that she could not carry the camera round the zoo. It was pointed out to them that there were lockers beside some of the cages in the zoo (at Locations 1, 2, 4, and 5 along the path), and that these lockers could be opened up and objects placed inside them. It was suggested, "You could leave the camera in one of these lockers and then Molly could get it when she's going to the zoo."

As with the search task, the experimenter then showed children that there was a path in the zoo that the doll must follow, with the experimenter tracing the path (from the child's right to left) with her finger. They were told that Molly must follow the path whenever she went to the zoo and she could only go round the zoo once and could not revisit locations. Children were asked to point to the start and end of the path to show that they had understood the experimenter. Children were then told, "Molly wants to take a picture of the kangaroo when she goes to the zoo." They were asked to point to the location of the kangaroo in the zoo, with all children giving the correct answer. The experimenter pointed out that there was no locker beside the kangaroo's cage, and so there was no place to leave the camera. It was proposed that, "You could leave the camera in one of the other lockers and then Molly could get it when she goes round the zoo. What locker could you leave it in so that Molly is still able to take a photo of the kangaroo?" If they did not answer, the experimenter prompted them. However, no location was suggested to children. As in the search task, children's answer was taken as the location of the locker that they pointed to.

Results

In both tasks, there were four locations that the child could possibly choose (Location 3, the kangaroo's cage, was not a legitimate response as there was no locker beside this cage). For each task, two of these were correct; in the search task, these were Locations 4 and 5, and in the planning task, these were Locations 1 and 2. Table 1 summarizes children's performance. Two-tailed binomial tests revealed that on the search task, neither 3-year-olds nor 4-year-olds performed at a level above chance ($p > .05$ for both year groups). Five-year-olds, however, performed significantly above chance on this task ($p < .01$). Further analyses were carried out to determine whether there were any differences in performance among the age groups. These showed that although there was no significant difference in performance between the 3- and 4-year-olds, 5-year-olds performed significantly better than both these groups, $\chi^2(1) = 11.59, p < .01$, and $\chi^2(1) = 10.00, p < .01$, respectively.

As in the search task, neither 3- nor 4-year-olds performed at a level exceeding chance in the planning task ($p > .05$ for both year groups), but a greater proportion of 5-year-olds passed this task than would have been expected by chance ($p < .05$). Unlike the search task, however, there were no significant differences in performance between any of the age groups. Indeed, as can been seen in Table 1, the number of 5-year-olds passing this task is very similar to the number of 4-year-olds, although only the 5-year-olds performed significantly above chance as a group. For each age group, further analyses examined whether performance was significantly better on one of the tasks. For both the 3-year-olds and the 5-year-olds, there was no difference in children's performance on the two tasks. However, significantly more 4-year-olds passed the planning task than the search task, $\chi^2(1) = 4.29, p < .05$.

Table 1

Number of Children in Each Age Group Who Passed or Failed as a Function of Age Group and Task

	Age group	Task	Pass	Fail
Experiment 1	3-year-olds	Search	11	19
		Planning	16	14
	4-year-olds	Search	12	18
		Planning	20	10
	5-year-olds	Search	24	6
		Planning	21	9
Experiment 2	4-year-olds	Search	16	14
		Planning	19	11
	5-year-olds	Search	24	6
		Planning	23	7
Experiment 3	4-year-olds	Search	23	7
		Planning	16	14
Experiment 4	4-year-olds	Search	24	6
		Planning	18	12
Experiment 5	4-year-olds	Planning	17	13
Experiment 6	4-year-olds	Search	15	15

Note. $n = 30$ for each age group in each task.

Additional analyses of children's choices of location were also conducted. Children's location responses were divided into correct and incorrect answers, and for each of these categories of responses, two-tailed binomial tests were carried out to examine whether one location was chosen significantly more frequently than the other. Table 2 shows the location choices for each age group; correct choices for each task are in bold. In the search task, a significant majority of both 3- and 4-year-olds chose Location 1 over Location 2 whenever they gave an incorrect answer ($p < .05$ for both age groups). However, there was no tendency for 5-year-olds to favor either of these two locations. Whenever children gave a correct answer, there was no significant difference between the frequency with which Locations 4 and 5 were picked.

In the planning task, both the 3- and the 5-year-olds chose Location 4 significantly more often than Location 5 whenever they answered incorrectly ($p < .05$).

Table 2
Distribution of Responses Across Locations as a Function of Age Group and Task

			Location			
	Age group	Task	1	2	4	5
Experiment 1	3-year-olds	Search	15	4	**4**	**7**
		Planning	**5**	**11**	13	1
	4-year-olds	Search	14	4	**4**	**8**
		Planning	**3**	**17**	7	3
	5-year-olds	Search	2	4	**11**	**13**
		Planning	**3**	**18**	8	1
Experiment 2	4-year-olds	Search	4	10	**7**	**9**
		Planning	**0**	**19**	11	0
	5-year-olds	Search	3	3	**11**	**13**
		Planning	**6**	**17**	6	1
Experiment 6	4-year-olds	Search	5	3	7[a]	**15**

Note. $n = 30$ for each age group in each task. Correct responses are given in bold.
[a] Due to the layout of the lockers in Experiment 6, these are in fact choices of Location 3 rather than 4, see Figure 1c.

There was no difference in 4-year-olds' choice of location for incorrect answers. Whenever children responded correctly, 4- and 5-year-olds picked Location 2 significantly more frequently than Location 1 ($p < .05$). In this case, though, 3-year-olds did not choose one location significantly more often than the other.

Discussion

The results of this experiment suggest that by 5 years, children can reason about novel event series both in the past and in the future. However, children had difficulties reasoning about novel event sequences before the age of 5, and this was true for event sequences located in both the past and the future. These general findings are consistent with the previous research of Povinelli et al. (1999) and McCormack and Hoerl (2005, 2007), although the event sequences were longer in the current experiment than the two-event sequences used in those studies. However, although 4-year-olds did not perform at levels significantly above chance on either task, they performed significantly better on the planning task than on the search task, performing at a level on the planning task that was very close to that of the 5-year-olds (only one additional child passed in the 5-year-old group compared to the 4-year-old group). This preliminary finding might suggest that there is an asymmetry in children's ability to make temporal–causal inferences concerning past and future event sequences, with 4-year-old children more able when thinking about the future than the past.

Such a suggestion might be thought to run contrary to suggestions that temporal cognition develops in tandem for the past and future. However, it is important to rule out a potentially simple explanation of this initial task difference that is suggested by the analyses of children's choices of location, specifically 3- and 4-year-olds' errors on the search task. Children of this age chose Location 1 more often than Location 2 whenever they gave an incorrect answer. This suggests that some children may have been attempting to engage in a serial search strategy instead of attempting to search logically: In other words, children may have tended by default to start to search at Location 1 with the intention of then searching sequentially at Locations 2, 4, and 5. However, because children were given only a single opportunity to choose a search location, they would have been unable to complete such a serial search. This interpretation of error patterns is consistent with the previous findings of Wellman et al. (1979; Experiment 2), who have suggested that young children may adopt such an unreflective strategy when it is relatively effortless to search. Thus, it is possible that performance on the search task was poorer in 4-year-olds than on the planning task because performance on the search task in part reflects a default to use an unproductive serial search strategy. In the planning task, no such preference for the first location was observed.

Experiment 2

Experiment 2 was designed as a replication of Experiment 1, but there were a few important changes to the procedure. The most pertinent, given the patterns of location choices in the previous experiment, was that the method of responding was manipulated to discourage any tendency in young children to search serially. In Experiment 2, children were given a single sticker with which to indicate their response, instead of being allowed to point to their answer. By giving children only one sticker, it was hoped that children would realize that they could only choose one location. Furthermore, if it is assumed that pointing to the first location (to initiate a serial search) is a prepotent response, then previous research would suggest that introducing an alternative method of indicating an answer might ensure that children give a more considered response (e.g., Hala & Russell, 2001). Experiment 2 also addressed the possibility that 3- and 4-year-old children's failure on the tasks could have been due to a memory failure rather than a failure to reason correctly. In particular, there are two important pieces of information that must be remembered in order to make the appropriate temporal–causal inferences. These are (a) the direction the doll went round the zoo or was going to go round the zoo (indicating an awareness of the order in which events happened/are to happen) and (b) the location where the camera had been used or was going to be used (the causally significant piece of information that determines where the child should look for or place the camera). Therefore, in Experiment 2, two memory control questions were included to ascertain whether this information was remembered.

Method

Participants

One hundred and twenty children took part in the task. Sixty 4-year-olds were tested, 30 on the search task (M = 53 months, range = 48–58 months) and 30 on the planning task (M = 53 months, range = 48–58 months). Similarly, thirty 5-year-olds were tested on the search task (M = 68 months, range = 60–71 months) and 30 on the planning task (M = 67 months, range = 60–71 months). There were 59 boys and 61 girls. These children were recruited from nurseries and schools local to the first author. They were primarily from working- and middle-class background. One hundred and seventeen children were of Caucasian descent, 2 were of Black descent, and 1 was of Asian descent. None of the children had previously taken part in Experiment 1. Children were tested individually in a quiet part of their school, and each was given a set of stickers for participating. Only children who gave the correct answer to both the memory control questions were included in the sample. An additional five children were tested who were subsequently excluded from analysis due to failure to answer one or both these questions correctly.

Materials

These were identical to Experiment 1, with the addition of small circular stickers. In both tasks, one sticker was given to participants and they had to use this to indicate their response.

Procedure

Search task. This was identical to the procedure used in Experiment 1 up until the point at which children were asked the test question. After the test question, children were given a small circular sticker and asked to place the sticker on the locker where they thought the camera might be in. They were told that they had only one sticker so could only look in one location. Before the experimenter looked into the chosen locker to see if the camera was there, but after children had made their choices, two memory control questions were asked. Children were asked what direction the doll had gone round the zoo and also what animal the doll had taken a photo of.

Planning task. This followed the procedure of Experiment 1 up until the end of the familiarization with the zoo. Instead of proceeding to the test question, children were given a brief distracter task, identical to that used in the search task. This made the planning task more comparable to the search task in that there was a short period of time when children's attention was not directly focused on the task. When children returned, they were asked the same test question as in Experiment 1, but they were now given a sticker and asked to place it on the locker that they thought the camera should be left in. After children had placed the sticker on one of the lockers, they were then asked the appropriate version of the two memory control questions.

In addition to these procedural changes, the direction in which the doll traveled/was to travel round the zoo was counterbalanced in this version of the task; for half the children, it was clockwise and for half anticlockwise.

Results

Initial analyses confirmed that there was no effect of the direction of travel in the task (clockwise or anticlockwise). This finding is consistent with the previous findings in the search literature (Haake & Somerville, 1985). Two-tailed binomial tests showed that 4-year-olds failed both the search task ($p > .05$) and the planning task ($p > .05$). Five-year-olds again performed significantly above chance on both tasks ($p < .01$). There was a significant difference between the two age groups in the search task, $\chi^2(1) = 4.8, p < .05$, but no difference between these age groups in the planning task, $\chi^2(1) = 1.27, p > .05$. There was now no difference in 4-year-olds' performance on the two tasks, $\chi^2(1) = .62, p > .05$. Unlike in Experiment 1, this age group did not do better on the planning task than on the search task. There was again no significant task difference for the 5-year-olds, $\chi^2(1) = 0.10, p > .05$.

On the search task, 4-year-olds now showed no significant preference for either Location 1 or Location 2 when their answer was incorrect (see Table 2; $p > .05$). Five-year-olds also showed no significant preference for one of the two locations when their answer was incorrect. Whenever a correct answer was given in the search task, neither 4- nor 5-year-olds chose one of the two locations significantly more than the other. In the planning task, however, there were significant differences in children's choice of location both when they gave a right and a wrong answer. Four- and 5-year-olds were more likely to choose Location 2 over Location 1 whenever they gave a correct response ($p = < .001$ and $p < .05$, respectively). Whenever an incorrect answer was given, 4-year-olds more often chose Location 4 than Location 5 ($p < .001$); this difference was not significant for the 5-year-olds.

Discussion

These results supported the main finding of the first experiment, which was that 5-year-olds were successful on both tasks. However, children younger than 5 years again had difficulties with both the search and the planning tasks. Unlike in Experiment 1, 4-year-olds did not find the planning task significantly easier than the search task. The main reason for the disappearance of the task effect appears to be the experimental modification introduced in the second experiment—the use of a single sticker as a mode of response. In Experiment 2, 4-year-olds showed no preference for choosing the first location in the search task, suggesting that this modification successfully discouraged children from using a serial search strategy. Thus, the results seem to suggest that the task difference for 4-year-olds found in Experiment 1 may have been due to the adoption of a particular strategy that may have somewhat depressed performance on the search task rather than any fundamental difference between the ability to reason about the past and the future. What the overall results of Experiment 2 show, though, is that children's failure of the search task in Experiment 1 was not solely due to errors resulting from a serial search strategy, as 4-year-olds still did not pass this task even when the response mode was modified. This is despite the fact that all the children included in the analyses answered the memory control questions correctly. This means that it cannot be that children's incorrect answers to test questions are due to a failure to remember these important bits of information. It is clear that knowledge of these pieces of information is not enough for successful performance; rather, an additional important cognitive skill is needed, which requires integrating these pieces of information.

Before turning to the next experiment, one important issue that needs to be considered is whether 3- and 4-year-olds' difficulties on the tasks in Experiments 1 and 2 were due to the fact that the object in question was a camera and the relevant event involved taking a photograph. Such an interpretation of children's difficulties might be raised by Zaitchik's (1990) influential research that has suggested that young children have some difficulties in fully grasping the representational nature of photographs. One reason for doubting such an interpretation is that subsequent research has suggested that 3-year-olds can pass Zaitchik's false photograph task if their attention is focused on the photograph itself (Perner, Leekam, Myers, Davis, & Odgers, 1998; Slaughter, 1998; though see Sabbagh, Moses, & Shiverick, 2006). In any case, our search and planning tasks placed minimal demands on the understanding of photographs and certainly less than required to pass the false photograph task. In particular, children were not required to grasp that photographs may fail to match current reality. The future task simply required children to know that a certain object (i.e., the camera) will be needed at a certain point in time. To interpret the test question, children did not need to know anything about the nature of photographs over and above the bare fact that the activity to be carried out in the future—the taking of the photograph—would involve the camera. The pretraining session provided a demonstration of this function of the camera. The search task did place an additional demand on children's understanding of photographs, in that the photograph taken of the kangaroo present during test was intended to serve as a reminder to children of when the camera had been used. In order for the photograph to act effectively as a cue, children needed to grasp that it was taken using the camera when the doll visited the kangaroo. Such a grasp has two elements: (a) basic knowledge that the photograph was taken by the camera (an analogous demand to that of the future task) and (b) an awareness of the correspondence between the object depicted in the photograph and the real object. A considerable body of research suggests that by 3 years, children have no difficulties grasping such photograph – object correspondences (e.g., DeLoache & Burns, 1994; Liben, 2003; Suddendorf, 2003).

Experiment 3

The results of the first two experiments suggested that the children younger than 5 years have difficulties making inferences about past and future event sequences. However, 3- and 4-year-olds' difficulties on the search task might be thought to contrast with the success of much younger children on tasks that to some extent share its structure. In double invisible displacement tasks, children have to search for an object that has been invisibly displaced behind one of two occluders that have been sequentially visited, with correct search involving ignoring irrelevant distracter occluders. Recent studies of invisible displacement in children have been carried out in a comparative context (Barth & Call, 2006; Collier-Baker & Suddendorf, 2006); taken together, the general conclusion that can be drawn from these studies is that 2-year-olds can pass at least some double invisible displacement tasks. Two-year-olds' success on such tasks is in marked contrast with the difficulties of 3- and 4-year-olds on our search task in Experiments 1 and 2 and forces a consideration of why our task is harder.

There are important differences between our search task and double invisible displacement. In many versions of the latter task, the order in which locations were visited is irrelevant, and indeed children seem to choose randomly between first and second visited locations (Collier-Baker & Suddendorf, 2006). However, there is in fact a more complex version of the task which is structurally more similar to our search task and that does seem to require a consideration of event order. Somerville and Capuani-Shumaker (1984) used a multiple-trial object search task in which children had to search for an object that was initially hidden in the experimenter's hand that they then tracked as it moved sequentially behind a series of four screens. In one version of the task, after the hand had visited the second screen, children were shown whether or not it still contained the object. Unlike in most invisible displacement tasks, the hand then visited the remaining two screens, with search being postponed until all four screens had been visited. Event order was important because correct search depended on considering whether the object was still in the experimenter's hand after it had visited the first two screens but before it had visited the second two screens (parallel to the consideration in our search task to the need to consider the fact that the camera had been used at a specific point in the event series). In fact, even 3- and 4-year-olds were by no means at ceiling on this task, particularly on second searches. However, 3-year-olds were above chance on at least some types of trials.

The important difference between such tasks and our task seems to be whether or not children are aware in advance of the event sequence that they will need to search for the object. In Somerville and Capuani-Shumaker's (1984) task, children were able to keep track of the object's trajectory as it moved through the various locations and sequentially update their model or models of the object's location as events unfolded, albeit being required to do so without actually being able to continuously see the object. The relevant distinction here is that between sequential updating one's model of the world as events unfold (which is sufficient for passing invisible displacement tasks) and retrospectively engaging in temporal–causal reasoning, a distinction that is described in more detail by McCormack and Hoerl (2005). In our task, children are required to retrospectively reconstruct and reason about the order in which certain events have occurred precisely because they had not been tracking the relevant object's location.

One way to put this point is to consider whether the tasks necessarily involve representations that have past tense content. Arguably, in single invisible displacement tasks, the content of children's representations can always be described as present tense. For example, when the object in the experimenter's hand moves behind the occluder, the child needs to form the representation "The object is in location X." When the hand emerges, they need not update their representation (because the hand is empty); they can then use this representation as the basis for object search. What is distinctive, and perhaps more demanding, about double invisible displacement tasks is that these representations must allow for the possibility that the object is in more than one location (i.e., "The object is in either X or Y"). However, the issue of whether the object must be represented as currently being in one of two possible locations is distinct from the issue of whether representations with a past tense content are needed for task success. Because in double invisible displacement, children can actively keep track of the object's location as events unfold, their representations need only involve present tense content because such representations can be updated online as the display changes. Thus, although task success may depend on children's being able to represent the object as currently being in either one of two places, it need not require retrospective reflection on the order of past events. However, in our search task, because children are not aware that they need to maintain a representation of the object's possible location as the event sequence unfolds, task success depends upon considering what has happened in the past and, in particular, the order in which the events have happened.

Although we have argued that our task is importantly different from that of Somerville and Capuani-Shumaker (1984), 3-year-olds' difficulties on their task contrast with the successful performance of younger children on Haake and Somerville's (1985) task that was structurally similar but had a shorter event sequence. In the latter task, there were only two occluders, one before and one after the point at which children saw whether or not the object was still in the experimenter's hand. Experiment 3 was designed to investigate whether a reduction in the number of events/locations would also aid young children's performance on our search task. In this experiment, the number of cages in the zoo was reduced from five to three (see Figure 1b), meaning that there was now one location before the kangaroo and one after it. All other aspects of the tasks remained identical to that of Experiment 2. As 5-year-olds had consistently proven to be successful on the original version of the tasks, we chose to only examine the abilities of 4-year-olds on this simplified task.

Method

Participants

Sixty 4-year-olds were tested, 30 on the search task (M = 56 months, range = 51–59 months) and 30 on the planning task (M = 55 months, range = 48–59 months). There were 34 girls and 26 boys. Five other children were tested but were excluded from the analysis as they answered one or more of the control questions incorrectly. All children were of Caucasian descent, except for one who was of Black descent and one who was of Asian descent. These were primarily of middle-class background. All the children attended schools close to the authors' university. Children were tested in a quiet area of the school and received a set of stickers for taking part.

Materials

The materials were identical to those used in Experiment 2, except that there were now only three cages used rather than five and, consequently, two lockers instead of four. The kangaroo was again always at the middle location, which was now Location 2 along the path.

Procedure

This was identical to that used in Experiment 2.

Results and Discussion

In the three-event/location version of the tasks, there was now only one right and wrong answer for each task. Two-tailed binomial analysis showed that 4-year-olds once again failed the planning task ($p >$ .05). However, 4-year-olds performed above chance on the search task ($p <$ .05). The difference in performance between the two tasks approached significance, $\chi^2(1) = 3.59$, $p = .06$. Thus, the results showed that 4-year-olds were able to pass a simplified version of the search task. It would appear that the reduction in the number of events/locations from five to three aided children's performance in this task. This modification of the task made no difference to children's performance on the planning task, however. Children of this age failed even this simplified version of the planning task.

Thus, the key finding of this experiment is that simplifying the event sequence had a different effect on the two tasks: improving performance on search but not on planning. As with Experiment 1, it is necessary to consider why there might be a disparity between the findings of the tasks assessing reasoning about past and future event sequences. One possible difference between the tasks may lie in the salience of event order information—information that is crucial for passing the task. In the search task, children have viewed the events in a specific sequence, and existing research on children's order memory would suggest that they should be able to remember this order (summarized in Friedman, 2003). In contrast, in the planning task, children are merely told about events that will occur in a specific sequence and may neglect order information in their reasoning even though they can answer memory control questions correctly.

Experiment 4

This experiment had two aims. First, we wished to replicate the findings of Experiment 3 in which children were successful on the simplified search but not the planning task. Second, we wanted to examine whether cueing children to focus on the information necessary to correctly answer the test questions would assist performance. Previous research using more complex planning tasks suggests that cues or external guidance can aid children's performance (e.g., Hudson & Fivush, 1991). Povinelli et al. (1999) also found that 5-year-old children could succeed on their temporal–causal reasoning task whenever the order in which the two causally relevant events occurred was emphasized, whereas without such cues, the group failed to perform at above-chance levels. However, McCormack and Hoerl (2005) found that the provision of cues did not improve 4-year-olds' performance on their temporal–casual reasoning task

In Experiments 2 and 3, children had been asked memory control questions after the test question. It was decided that asking such control questions might act as sufficient cues, if they were asked before the test question and framed in such a way that the child was made to reflect on this information. Although 4-year-olds in Experiments 2 and 3 could remember the appropriate pieces of information, it does not necessarily follow that they attended to them during the task. It was hoped that questioning children about this information immediately prior to being asked the test question would make these pieces of information particularly salient.

Method

Participants

Participants were sixty 4-year-olds, 30 tested on the search task ($M = 55$ months, range = 50–59 months) and 30 on the planning task ($M = 54$ months, range = 48–59 months). There were 31 boys and 29 girls. These children attended nursery schools local to the authors' university and were tested individually in a quiet room in the nursery. All were of Caucasian descent. Children received a set of stickers for taking part.

Materials

These were identical to those used in Experiment 3.

Procedure

The tasks followed the procedure of Experiment 3 except that the memory control questions were now asked before the test question instead of after it. In the search task, before the test question, the experiment said: "Let's now think about where the camera must be. Do you remember what animal Molly took a photo of? That's right, look here's the photo of the kangaroo. And what way did Molly go round the zoo? That's right." In the planning task, before asking the test question, the experimenter said: "Let's think about where we should put the camera so that Molly can take a photo of the kangaroo. What way does Molly have to go round the zoo? That's right. And what animal does Molly want to take a photo of? That's right." The order in which the control questions were asked was counterbalanced.

Results and Discussion

Initial analyses showed that the order in which the memory control cues were asked did not influence children's performance. Two-tailed binomial tests showed that 4-year-olds again passed the search task ($p < .01$). However, as a group, they failed the planning task ($p > .05$). Analyses comparing these results with those of Experiment 3 showed that there was no significant improvement in 4-year-olds' performance in either the search task, $\chi^2(1) = 0.1$, $p > .05$ or the planning task, $\chi^2(1) = 0.27$, $p > .05$. The results of this study closely replicated those of Experiment 3, in that 4-year-olds passed a simplified version of the search task but not the planning task. Although children passed the search task, they were not at ceiling as a group and performed at a very similar level to the 4-year-olds in Experiment 3. Thus, making children explicitly attend to the relevant information did not improve children's ability on either task compared to when they had to spontaneously recruit that information. In this sense, these findings are consistent with those of McCormack and Hoerl (2005). However, the finding in Experiments 3 and 4 that 4-year-olds can pass the simplified search but not the planning task stands in direct contrast to the findings of Experiments 1 and 2 in which children of this age group failed both tasks.

The consistency of the findings of Experiments 3 and 4 suggests a reliable difference between children's performance on the search and planning tasks on this simplified version of the tasks. This difference suggests that although 4-year-olds may be able under some circumstances to make temporal–causal inferences for past events, they have more difficulties when the events in question are in the future. The findings of the first three experiments raise two important issues: (a) why do children younger than 5 years have difficulties with the planning but not the search task and (b) why do 4-year-old children find the three-event/location search task easier than the five-event/location task? These two issues were explored in Experiments 5 and 6, respectively.

Experiment 5

The previous studies have shown that 4-year-olds fail the planning task even when the event sequence has been simplified and when they have been specifically cued to the relevant pieces of information. One obvious difference between the search and planning tasks lies in the extent to which imagination is required in both. In the search task, children have observed the sequence of past events that they must reason about, whereas in the planning task, children must imagine and reason about never-before-experienced events taking place in an unknown future. Harris (2000) has convincingly argued that research on children's play and pretense suggests that the causal implications of imagined actions or events are grasped from an early age, and others have suggested that children may possess the imaginative abilities necessary for simple planning relatively early in development (Haith, 1997). However, play and pretense usually involve imagining versions of events or actions that are in many ways already familiar to children. Furthermore, imagining the causal consequences of a single imagined act is likely to be more developmentally primitive than grasping the significance of the temporal order of a sequence of imagined events. The current task was designed so that it involved a novel event sequence, and it required children to mentally construct a sequence of future events while bearing in mind a novel goal.

In Experiment 5, we examined whether it might be possible to boost children's performance by making this imaginative project easier. Previous research on children's planning across a variety of tasks has suggested that provision of goal state information may aid performance (Bauer et al., 1999; Shapiro & Hudson, 2004; Welsh, 1991). In Experiment 5, children were actually provided with a concrete example of the goal in the task—namely, to obtain a photograph of the kangaroo. In the search task, a photograph of the kangaroo was present throughout the testing phase. In Experiment 5, the test procedure was altered so that the planning task resembled the search task in this respect. Children were told that another doll had already been to the zoo prior to Molly and taken a photo of the kangaroo, which they were then shown. They were then informed that Molly also wanted to take a photo of the kangaroo. This photo of the kangaroo remained present throughout the task. If 4-year-old children perform successfully on this version of the planning task, it might suggest that their failure of the task in Experiments 1–3 was due to the particular demands of imagining a novel goal state when thinking about the future in contrast to reasoning about a known past.

Method

Participants

Thirty 4-year-olds ($M = 53$ months, range = 48–59 months) completed this version of the planning task. There were 13 boys and 17 girls. Three additional children were excluded from the analysis because they failed to answer one or more of the control questions correctly. Participants were mostly from working- and middle-class backgrounds and all were of Caucasian descent. All children attended a nursery school local to the authors' university. Children were given a set of stickers for participating. None of these children had taken part in previous experiments.

Materials

These included the same materials that had been used in the planning task in Experiments 3 and 4. In addition, there was another doll, Molly's sister Rosie. This was of similar size and shape to the first doll. A green felt rucksack (of the type used in the previous search tasks) was also used in the introductory phase of the task. Inside this was placed another Polaroid camera identical to the one used by Molly. This bag also contained a photo of the kangaroo that had been taken by the Polaroid camera.

Procedure

After the initial familiarization phase, children were told that Molly had a sister, Rosie, who had already been to the zoo. They were shown this doll, which had a bag on its back containing the Polaroid camera and the photo of the kangaroo. The camera was taken out of this doll's bag, and children were told that Rosie also had a camera, which she took to the zoo with her. They were shown that this camera was the same as Molly's. They were informed that Rosie had taken a photo of the kangaroo when she had been at the zoo, and they were shown the photo. This was then placed in front of the kangaroo's cage, directly facing children, where it stayed for the remainder of the task. Children's attention was then drawn to Rosie's bag, and the experimenter told them that Rosie had carried her camera in this bag. They were then told that Molly did not have a bag with her and so could not carry the camera round the zoo. From this point onward, the procedure was identical to that used in Experiment 3.

Results and Discussion

As can be seen in Table 1, the modification of the task did not improve 4-year-olds' performance. This group again failed to perform at a level significantly above chance ($p > .05$, two-tailed binomial). Thus, children's failure of this task appears to be a robust phenomenon. Children of this age could still not make the correct inference even when they were provided with a visually available example of the goal state in the task. This provides some indication that younger children's problems on the planning task may not be due to imagining a novel goal state. Children's persistent difficulties on this task contrast with their good performance on the simplified search task in

Experiments 3 and 4 and force a consideration of the differences between reasoning about the past and the future.

Experiment 6

Experiments 3 and 4 showed that 4-year-olds can pass a simplified version of the search task involving a reduced number of events/locations, whereas they are unsuccessful on the five-event/location version. There are at least two clearly separable explanations of why reducing the number of events/locations might improve children's performance. One of these is to do with the presence of multiple possibilities. In the task used in Experiments 1 and 2, there were two possible right answers: The camera could have been lost at either the fourth or fifth location. Both of these were equally likely to be the true location of the camera. Children were only allowed to search in one location, however, and so had to choose between them. Being forced to choose between them may have adversely affected younger children's performance because of the inherent ambiguity over which location is correct. The second possible type of explanation of the improved performance in the search task in Experiments 3 and 4 relates directly to the actual number of events and locations used in the task rather than to the number of possible correct answers. Arguably, the array is more visuospatially complex when there are more locations, and there are greater number of events to represent. Reducing the number of events/locations in itself may mean that the task has a reduced processing load.

The first explanation is quite different from the second and implicates a fundamental problem with reasoning when there are multiple possibilities rather than an excessive demand on cognitive processes. Recent empirical research has shown that younger children consider multiple possibilities or ambiguity differently from older children and adults (Beck, Robinson, Carroll, & Apperly, 2006; Robinson, Rowley, Beck, Carroll, & Apperly, 2006; though see Sophian & Somerville, 1988). This research indicates that 4-year-olds may not realize that multiple possibilities exist, or consider them in their reasoning, rather than the presence of multiple possibilities interfering with their reasoning skills. However, it is possible that problems reasoning in a task in which there is more than one possible correct answer produced difficulties for children in the search task used in Experiments 1 and 2. In Experiment 6, 4-year-olds were given a five-event/location search task in which there was only one correct answer. In the new configuration, there were five locations/events, as in Experiments 1 and 2, but three of these were before the kangaroo and only one after it (Figure 1c). Thus, as in Experiments 3 and 4, there was a single location in which the camera could logically be. If 4-year-olds were successful on this task, then this would suggest that it was something to do with having more than one correct location (i.e., difficulties dealing with multiple possibilities) that disrupted children's ability to make the appropriate inferences in Experiments 1 and 2.

Method

Participants

Thirty 4-year-olds were tested on this version of the search task ($M = 53$ months, range = 49–58 months). There were 13 boys and 17 girls. Two additional children were tested but were not included in the analysis as they gave the wrong answer to one or more of the control questions. All children were of Caucasian descent and were from mainly working- and middle-class backgrounds. Participants were drawn from nursery schools local to the first author, and testing took place in a quiet part of the nursery in a room suitable for testing purposes. Children received a set of stickers for taking part. None of the children had taken part in any of the previous experiments.

Materials

The materials that had been used in the original version of the search task in Experiment 2 were again utilized in this study, arranged as illustrated in Figure 1c.

Procedure

The procedure was identical to that of Experiment 2.

Results and Discussion

In this study, Locations 1, 2, and 3 were scored as incorrect, whereas Location 5 was correct; Location 4 was the kangaroo's cage and thus not a possible choice. Children's location responses are shown in Table 2, with correct answers shown in bold. It can be seen that exactly half the group chose the correct location, whereas the other half chose one of the other three locations. If a binomial test proportion of .5 is used, with children's answers categorized dichotomously as correct and incorrect, then the group can be interpreted as failing to choose the correct location significantly more often than would be predicted by chance ($p = 1.00$). However, there are four locations, and if it is assumed that each has a one in four chance of being chosen, then a binomial test proportion of .25 should be employed (this issue did not arise in previous analyses because exactly half of the locations were always correct or incorrect). Such an analysis showed that children chose Location 5 (the correct answer) significantly more frequently than any other location ($p < .01$). Thus, even though a majority of children did not answer correctly, it might be argued that nevertheless the group as a whole was not responding at random. In order to interpret this nonrandom performance, we compared children's location choices in this task with those in the search task in Experiment 2. The rationale for this comparison was that if it is the case that children's nonrandom performance in this experiment should be interpreted in terms of task competence, then their performance should differ significantly from that in the version of the task used in Experiment 2 in which the locations had a different status. The analysis showed that location choices in this task did not in fact differ significantly from those in Experiment 2, $\chi^2(3) = 5.38$, $p \geq .05$. Thus, despite there being only one correct answer, this age group still failed the task. This suggests that it was not the ambiguity created by having more than one possible correct location that was responsible for the difference in performance between the three- and five-location search tasks. Rather, 4-year-olds' success on the simplified version of the task appears to be a direct result of reducing the number of events/locations.

General Discussion

The aim of this research was to investigate a particular type of mature temporal thinking—temporal–causal reasoning—that involves reasoning about the causal significance of the temporal order of events (McCormack & Hoerl, 2005, 2007; Povinelli et al., 1999). The search and planning tasks allowed us to examine the development of this type of thinking for both past and future sequences of events. Taken as a whole, the findings suggest that by 5 years, children can reason about the causal relations between temporally ordered events for both the past and the future. By contrast, 3- and 4-year-olds failed the original versions of both the search and the planning tasks, even when the response mode was changed to discourage unproductive search strategies (Experiments 1 and 2). Memory checks ensured that this failure was not due to a difficulty remembering the necessary pieces of information to make the appropriate inferences (Experiment 2). These initial results are consistent with the previous research in this area (McCormack & Hoerl, 2005, 2007; Povinelli et al. 1999), despite very different methodologies.

Search

Although 4-year-olds were not above chance on the search task in Experiments 1 and 2, they were successful on a simplified version of the search task in which there were only three events in the sequence and only one possible correct answer (Experiments 3 and 4). This finding raises two closely related questions: (a) why does simplifying the task in this way lead to improved performance and (b) what do the findings tell us about the development of this type of reasoning? The first question was addressed in Experiment 6, in which the number of events/locations was identical to that used in the first two experiments (i.e., five), but the array was altered so that there was only one possible correct response. Four-year-olds were unsuccessful in this version of the task, which suggests that reducing the overall number of events/locations in Experiments 3 and 4 in itself improved performance on our search task rather than the elimination of the possibility of more than one correct response. This interpretation of the findings is consistent with the findings of Collier-Baker and Suddendorf's (2006) recent study of double invisible displacement. They found that under some conditions, 2-year-olds could appropriately direct their second searches if they failed to find the object on a first search, indicating that under some circumstances, even much younger children find the presence of more than one correct response unproblematic.

The second question forces a consideration of whether we should conclude that children younger than 5 years are capable of making appropriate temporal–causal inferences when the task is sufficiently simple. Such a conclusion could be interpreted as inconsistent with the findings of Povinelli et al. (1999) and McCormack and Hoerl (2005, 2007), as children of this age had consistent problems in their tasks. We have argued that the nature of our task makes it possible to straightforwardly interpret it as assessing this type of reasoning, in contrast to various earlier studies of children's search behaviors that were not designed for that purpose (e.g., those of Anooshian et al., 1982; Haake et al., 1980; Wellman et al., 1979). For example, we have ensured that children's searches did not simply reflect basic associative processes or that failure was not due to memory problems. The spatial layout, even in Experiments 1 and 2, was also simple. However, the fact that 4-year-olds passed the task only when the spatial layout was further simplified (Experiments 3 and 4) might in itself suggest a way of reconciling our findings with those of Povinelli et al. and McCormack and Hoerl (2005, 2007). Specifically, perhaps children succeeded on some versions of the task precisely because it had a spatial component.

In designing the tasks, it was hoped that mirroring temporal order in spatial order would support children's sequence memory and ensure that they did not fail the task because they had forgotten the order in which events had occurred. It may be the case that making temporal order concrete in this way—by spatializing it and making it visually available—supports 4-year-olds' inferences concerning temporal order. In particular, if the spatial layout is very simple, as in Experiments 3 and 4, then the fact that temporal order is in some sense visually available may assist 4-year-olds in making an appropriate inference. In summary, the findings suggest that there may be some circumstances under which 4-year-olds are able to make temporal–causal inferences about past events and these may include circumstances in which temporal order is mirrored in a simple spatial display.

Planning

Five-year-olds' consistent success on the planning task in Experiments 1 and 2 demonstrates that by this age, children can construct a plan concerning a completely novel event sequence, taking into account the order in which events will occur. However, 4-year-olds consistently failed the planning task even when the number of events in the sequence was reduced and even when they were asked to recall the pieces of information necessary to make the correct inference immediately before the test question (Experiment 4). As they stand, these findings are consistent with the previous research into young children's planning abilities, which has found a marked improvement in planning ability between the ages of 3 and 5 (Hudson et al., 1997; Wellman et al., 1985; Welsh, 1991).

Four-year-olds' passed neither the search nor planning tasks in Experiments 1 and 2. However, this age group passed the simplified search but not the planning task in Experiments 3 and 4. One possible difference between the two tasks is that reasoning about a completely novel sequence of future events may place greater demands on imagination than reasoning about a sequence of past events that has already been observed (see also Hudson et al., 1997). In other words, perhaps children's difficulties with planning on the task were primarily problems with generative imagination rather than reasoning per se. In Experiment 5, children were provided with an example of the goal state in the planning task in an attempt to make this imaginative project easier. This sort of visual support had been present during the test phase of the search task but not in the versions of the planning task used in Experiments 1–4. However, its provision in Experiment 5 did not lead to improved performance. Of course, it is possible to argue that there may be additional, more effective, ways in which the imaginative demands of the task might be reduced: For example, we could have actually shown the other doll travel round the zoo and take a photograph of the kangaroo. It may be the case that the fact that the event sequence was entirely novel made planning particularly difficult in this task. In designing future tasks, it may prove useful to distinguish between the imaginative demands of a planning task and the reasoning processes that it involves.

The Development of Temporal Cognition

We finish by turning to a consideration of the relationships between the type of ability we have studied and two other temporal abilities that have been the focus of research: MTT and temporal decentering.

MTT

The aim of our studies has been to examine the development of a particular type of temporal cognition that may be involved in reasoning about both past and future event sequences. There are at least three similarities between the type of temporal cognition we are concerned with and that which is the focus of researchers working on the development of MTT. First, in designing our tasks, we have attempted to ensure that children need to detach from the here and now and consider the implications of either what has happened at a specific point in the past or what will happen at a point in the future. Similarly, researchers working on MTT are trying to capture a type of cognition that involves mentally detaching from current events or concerns. Second, we are focusing on a cognitive ability that goes beyond the ability to bring to bear knowledge of familiar event sequences. Thus, our tasks require children to reason about entirely novel unfamiliar event sequences. Researchers working on MTT are also trying to pinpoint a type of cognition concerning specific past or future events (hence the term *episodic future thinking*; Atance & O'Neill, 2001a) and contrast this type of cognition with the ability to draw on semantic memory or on script-like memory (Atance & Meltzoff, 2005; Suddendorf & Busby, 2003). Third, because of this characteristic, both temporal – causal reasoning and MTT are thought to be flexible and insightful types of cognition—for example, the type of planning for the future that they involve is not merely a matter of bringing to bear preexisting knowledge about what usually happens in such a situation or using a preestablished routine or strategy. Indeed, one way to think about the relationship between the two cognitive abilities is that MTT may deliver representations of specific past and future events, but frequently temporal – causal reasoning may be required to make use of the information given in such representations through a consideration of the causal connection between such events and other relevant events.

However, there are also important differences between the way that we conceptualize temporal – causal reasoning and how MTT has been characterized. It has been argued there is a distinctive phenomenology associated with the mental preexperiencing of past events and the imaginative preexperiencing of future ones that singles out MTT from other types of cognition concerning the past and the future (Suddendorf & Busby, 2003; Tulving, 2001). Furthermore, MTT is thought to be an essentially self-conscious activity involving one's personal past and future (Tulving, 2005). In contrast, we have defined temporal – causal reasoning without reference to subjective experience, and our experimental approach does not allow us to examine whether young children's thought about the past and future possesses the relevant distinctive phenomenology. For this reason, we would not wish to argue that the tasks used in our studies necessarily involve MTT. With regard to the involvement of self-consciousness, it is clear that the tasks do not involve either making plans that involve the self carrying out an action (Atance & O'Neill, 2001a, 2005; Suddendorf & Busby, 2005) or reflecting on one's personal past (Nelson, 2001; Perner & Ruffman, 1995). Clearly, the same basic sort of reasoning ability would be required if our tasks were set in contexts in which the actor were the child rather than a doll character. However, we do not know if the developmental trends and task effects that we have found would be replicated under such circumstances. Exploring such an issue is an important extension of this work.

Although our studies have clearly not addressed issues concerning subjective experience, there may still be conceptual connections to be drawn between temporal – causal reasoning and both self-consciousness and the phenomenology of thought about other times. For example, it has been argued elsewhere that possession of a mature concept of time may involve a grasp of the causal connections between events that occur at different times (Hoerl & McCormack, 2005; see also Martin, 2001), and it may be that possession of such a concept is a necessary developmental prerequisite for certain types of subjective experiences (McCormack & Hoerl, 2001). Furthermore, it may be the case that children initially acquire such a concept through reflection on the causal connections between events in their own personal past (Hoerl & McCormack, 2005; McCormack & Hoerl, 2001; see also Welch-Ross, 2001). Thus, although the sort of temporal – causal reasoning measured in our tasks is not a self-conscious activity, it may have its developmental origins in self-conscious reflection.

Temporal Decentering

Weist (1986, 1989) has suggested that the period between 3 and 5 years involves the emergence of temporal decentering, which he conceived of as the ability to think about events and the temporal relationships between them from a temporal perspective other than the present (see also Cromer, 1971, Harner, 1982; McCormack & Hoerl, 1999). Temporal decentering is viewed as a fundamental form of mature temporal cognition that applies to both the past and the future. Weist's account is based on a consideration of the ways in which children's temporal language changes during this developmental period. In some languages, children between 3 and 5 years begin to use complex tenses such as the past perfect tense, for example, "I had made the cake." Drawing on Reichenbach's (1947) analysis of tense, Weist suggested that use of such tensed forms indicates that children are able to flexibly coordinate at least three points in time—in the earlier example, the present time of utterance, the time at which the cake was made, and a subsequent time at which the cake had already been made. He argued that the coordination of three separate points in time is a hallmark of flexible temporal decentering. Attempts to examine such abilities experimentally have looked at whether children can match tensed sentences to appropriate points in an illustrated event sequence (Cromer, 1971; Harner, 1980).

Our tasks deliberately placed minimal demands on temporal language. However, arguably they assess a similar ability to that described by Weist as temporal decentering because they involve children having to consider the relationship between three points in time: the current time, the time at which the photograph has been or will be taken, and the time at which the camera was lost or will be collected. Weist (1989) has also argued that between 3 and 5 years, children begin to properly master the terms *before* and *after* (Harner, 1980; Trosborg, 1982), and that such mastery is further evidence of the development of flexible temporal decentering and mature temporal cognition. Although the terms *before* and *after* were not used in our tasks, successful performance depended on children in the search task realizing that they need to search in locations visited after the camera had been used and children in the planning task realizing that they needed to put the camera in a location to be visited before the camera was required. It would be interesting to explore whether success on the tasks is related to children's comprehension of temporal terms such as *before* and *after*, and more generally to children's mastery of temporal language.

Conclusions

Theorists describing cognitive abilities such as MTT have suggested that thinking about the past and the future recruit a common cognitive skill. We have argued that this is true, but we have characterized such a skill as involving a certain type of reasoning ability rather than in terms of a certain type of consciousness. Although the requirements of the two temporal – causal reasoning tasks used in these studies were not identical, a strength of the experimental method is that it provides a context in which differences between reasoning about the past and the future could be discovered and explored. Our results suggest that although thinking about the past and future both require an appreciation of the causal significance of the temporal order of events, there may be additional requirements specific to thinking about the future. Furthermore, with regard to both the past and future and in line with previous findings and theoretical claims (Atance & O'Neill, 2005; Friedman, 2005; McCormack & Hoerl, 2005, 2007; Povinelli et al., 1999; Suddendorf & Busby, 2005), we have found developmental changes that suggest a fundamental change in temporal cognition between 3 and 5 years.

References

Anooshian, L. J., Hartman, S. R., & Scharf, J. S. (1982). Determinants of young children's search strategy in a large-scale environment. *Developmental Psychology, 18*, 608 – 616.

Atance, C. M., & Meltzoff, A. N. (2005). My future self: Young children's ability to anticipate and explain future states. *Cognitive Development, 20*, 341 – 361.

Atance, C. M., & O'Neill, D. K. (2001a). Episodic future thinking. *Trends in Cognitive Sciences, 5*, 533 – 539.

Atance, C. M., & O'Neill, D. K. (2001b). Planning in 3-year-olds: A reflection of the future self? In C. Moore & K. Lemmon (Eds.), *The self in time. Developmental perspectives* (pp. 121 – 140). Mahwah, NJ: Erlbaum.

Atance, C. M., & O'Neill, D. K. (2005). The development of episodic future thinking. *Learning and Motivation, 26*, 126 – 144.

Barth, J., & Call, J. (2006). Tracking the displacement of objects: A series of tasks with great apes (Pan troglodytes, Pan paniscus, Gorilla gorilla, and Pongo pygmaeus) and young children (Homo sapiens). *Journal of Experimental Psychology: Animal Behavior Processes, 32*, 239 – 252.

Bauer, P. J., & Mandler, J. M. (1989). One thing follows another: Effects of temporal structure on 1- to 2-year-olds' recall of events. *Developmental Psychology, 25*, 197 – 206.

Bauer, P. J., Schwade, J. A., Wewerka, S. S., & Delaney, K. (1999). Planning ahead: Goal-directed problem solving by two-year-olds. *Developmental Psychology, 35*, 1321 – 1337.

Beck, S. R., Robinson, E. J., Carroll, D. J., & Apperly, I. A. (2006). Children's thinking about counterfactuals and future hypotheticals as possibilities. *Child Development, 77*, 413 – 426.

Benson, J. B. (1997). The development of planning: It's about time. In S. L. Friedman & E. K. Scholnick (Eds.), *The developmental psychology of planning: Why, how, and when do we plan?* (pp. 43 – 75). Mahwah, NJ: Erlbaum.

Busby, J., & Suddendorf, T. (2005). Recalling yesterday and predicting tomorrow. *Cognitive Development, 20*, 362 – 372.

Carlson, S. M., Moses, L. J., & Claxton, L. J. (2004). Individual differences in executive functioning and theory of mind: An investigation of inhibitory control and planning ability. *Journal of Experimental Child Psychology, 87*, 299 – 319.

Clayton, N. S., Bussey, T. J. , & Dickinson, A. (2003). Can animals recall the past and plan for the future? *Nature Reviews Neuroscience, 4*, 685 – 691.

Collier-Baker, E., & Suddendorf, T. (2006). Do chimpanzees (Pan troglodytes) and 2-year-old children (Homo sapiens) understand double invisible displacement? *Journal of Comparative Psychology, 120*, 89 – 97.

Corrigan, R., & Denton, P. (1996). Causal understanding as a developmental primitive. *Developmental Review, 16*, 162 – 202.

Cromer, R. F. (1971). The development of the ability to decentre in time. *British Journal of Psychology, 62*, 353 – 365.

DeLoache, J. S., & Burns, N. M. (1994). Early understanding of the representational function of pictures. *Cognition, 52*, 83 – 110.

Fivush, R., & Mandler, J. M. (1985). Developmental changes in the understanding of temporal sequence. *Child Development, 56*, 1437 – 1446.

Friedman, W. J. (1977). Development of children's understanding of cyclic aspects of time. *Child Development, 48*, 1593 – 1599.

Friedman, W. J. (2003). The development of children's understanding of the past and the future. *Advances in Child Development and Behavior, 31*, 229 – 269.

Friedman, W. J. (2005). Developmental and cognitive perspectives on humans' sense of the times of past and future events. *Learning and Motivation*, *36*, 145–158.

Gopnik, A., Sobel, D. M., Schulz, L. E., & Glymour, C. (2001). Causal learning mechanisms in very young children: Two, three, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental Psychology*, *37*, 620–629.

Goswami, U. (1998). *Cognition in children*. Hove, UK: Psychology Press.

Haake, R. J., & Somerville, S. C. (1985). Development of logical skills in infancy. *Developmental Psychology*, *21*, 176–186.

Haake, R. J., Somerville, S. C., & Wellman, H. M. (1980). Logical ability of young children in searching a large-scale environment. *Child Development*, *51*, 1299–1302.

Haith, M. M. (1997). The development of future thinking as essential for the emergence of skill in planning. In S. L. Friedman & E. K. Scholnick (Eds.), *The developmental psychology of planning: Why, how, and when do we plan?* (pp. 25–42). Mahwah, NJ: Erlbaum.

Hala, S., & Russell, J. (2001). Executive control within strategic deception: A window on early cognitive development. *Journal of Experimental Child Psychology*, *80*, 112–141.

Harner, L. (1980). Children talk about the time and aspect of actions. *Child Development*, *52*, 498–506.

Harner, L. (1982). Talking about the past and the future. In W. J. Friedman (Ed.), *The developmental psychology of time* (pp. 141–169). New York: Academic Press.

Harris, P. L. (2000). *The work of the imagination*. Oxford, UK: Blackwell.

Hassabis, D., Kumaran, D., Vann, S. D., & Maguire, E. A. (2007). Patients with hippocampal amnesia cannot imagine new experiences. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 1726–1731.

Hoerl, C., & McCormack, T. (2005). Joint reminiscing as joint attention to the past. In N. Eilan, C. Hoerl, T. McCormack, & J. Roessler (Eds.), *Joint attention: Communication and other minds* (pp. 260–286). Oxford, UK: Oxford University Press.

Hudson, J. A. (2002). "Do we know what we're going to do this summer?" Mothers talk to young children about future events. *Journal of Cognition and Development*, *3*, 49–71.

Hudson, J. A. (2006). The development of future time concepts through mother-child conversation. *Merrill-Palmer Quarterly*, *52*, 70–95.

Hudson, J. A., & Fivush, R. (1991). Planning in the preschool years: The emergence of plans from general event knowledge. *Cognitive Development*, *6*, 393–415.

Hudson, J. A., Sosa, B., & Shapiro, L. R. (1997). Scripts and plans: The development of preschool children's event knowledge and event planning. In S. L. Friedman & E. K. Scholnick (Eds.), *The developmental psychology of planning: Why, how, and when do we plan?* (pp. 77–102). Mahwah, NJ: Erlbaum.

Klahr, D., & Robinson, M. (1981). Formal assessment of problem-solving and planning processes in preschool children. *Cognitive Psychology*, *13*, 113–148.

Liben, L. (2003). Beyond point and shoot: Children's developing understanding of photographs as spatial and expressive representations. *Advances in Child Development and Behaviour*, *31*, 1–42.

Luciana, M, & Nelson, C.A. (1998). The functional emergence of prefrontally-guided working memory systems in four- to eight-year-old children. *Neuropsychologia*, *36*, 273–293.

Martin, M. G. F. (2001). Out of the past: Episodic memory as retained acquaintance. In C. Hoerl & T. McCormack (Eds.), *Time and memory: Issues in philosophy and psychology* (pp. 257–284). Oxford, UK: Oxford University Press.

McCormack, T., & Hoerl, C. (1999). Memory and temporal perspective: The role of temporal frameworks in memory development. *Developmental Review*, *19*, 154–182.

McCormack, T., & Hoerl, C. (2001). The child in time: Temporal concepts and self-consciousness in the development of episodic memory. In C. Moore & K. Lemmon (Eds.), *The self in time* (pp. 203–227). Mahwah, NJ: Erlbaum.

McCormack, T., & Hoerl, C. (2005). Children's reasoning about the causal significance of the temporal order of events. *Developmental Psychology*, *41*, 54–63.

McCormack, T., & Hoerl, C. (2007). Young children's reasoning about the order of past events. *Journal of Experimental Child Psychology*, *98*, 168–183.

Nelson, K. (1996). *Language in cognitive development: Emergence of the mediated mind*. Cambridge, UK: Cambridge University Press.

Nelson, K. (2001). Language and the self: From the "experiencing I" to the "continuing me". In C. Moore & K. Lemmon (Eds.), *The self in time: Developmental perspectives* (pp. 15–33). Mahwah, NJ: Erlbaum.

Perner, J., Leekam, S. R., Myers, D., Davis, S., & Odgers, E. (1988). Misrepresentation and referential confusion: Children's difficulty with false beliefs and outdated photographs. Retrieved July 16, 2007, from http://cogprints.org/708/

Perner, J., & Ruffman, T. (1995). Episodic memory and autonoetic consciousness: Developmental evidence and a theory of childhood amnesia. *Journal of Experimental Child Psychology*, *59*, 516–548.

Povinelli, D. J. (2001). The self: Elevated in consciousness and extended in time. In C. Moore & K. Lemmon (Eds.), *The self in time: Developmental perspectives* (pp. 75–95). Mahwah, NJ: Erlbaum.

Povinelli, D. J., Landry, A. M., Theall, L. A., Clark, B. R., & Castille, C. M. (1999). Development of young children's understanding that the recent past is causally bound to the present. *Developmental Psychology*, *35*, 1426–1439.

Raby, C. R., Alexis, D. M., Dickinson, A., & Clayton, N. S. (2007). Planning for the future by Western Scrub-Jays. *Nature*, *445*, 919–921.

Reichenbach, H. (1947). *Elements of symbolic logic*. New York: Macmillan.

Robinson, E. J., Rowley, M., Beck, S. R., Carroll, D. J., & Apperly, I. A. (2006). Children's sensitivity to their own relative ignorance: Handling of possibilities under conditions of epistemic and physical uncertainty. *Child Development*, *77*, 1642–1655.

Sabbagh, M. A., Moses, L. J., & Shiverick, S. (2006). Executive functioning and preschoolers' understanding of false beliefs, false photographs, and false signs. *Child Development*, *77*, 1034–1049.

Scholnick, E. K., Friedman, S. L., & Wallner-Allen, K. E. (1997). What do they really measure? A comparative analysis of planning tasks. In S. L. Friedman & E. K. Scholnick (Eds.), *The developmental psychology of planning: Why, how, and when do we plan?* (pp. 127–156). Mahwah, NJ: Erlbaum.

Shapiro, L. R., & Hudson, J. A. (2004). Effects of internal and external supports on preschool children's event planning. *Applied Developmental Psychology*, *25*, 29–73.

Slaughter, V. (1998). Children's understanding of pictorial and mental representations. *Child Development*, *69*, 321–332.

Somerville, S. C., & Capuani-Shumaker, A. (1984). Logical searches of young children in hiding and finding tasks. *British Journal of Developmental Psychology*, *2*, 315–328.

Sophian, C., & Somerville, S. C. (1988). Early developments in logical reasoning: Considering alternative possibilities. *Cognitive Development*, *3*, 183–222.

Suddendorf, T. (2003). Early representational insight: Twenty-four-month-olds can use a photo to find an object in the world. *Child Development*, *74*, 896–904.

Suddendorf, T., & Busby, J. (2003). Mental time travel in animals? *Trends in Cognitive Science*, *7*, 391–396.

Suddendorf, T., & Busby, J. (2005). Making decisions with the future in mind: Developmental and comparative identification of mental time travel. *Learning and Motivation*, *36*, 110–125.

Szpunar, K. K., Watson, J. M., & McDermott, K. B. (2007). Neural substrates of imagining the future. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 642–647.

Trosborg, A. (1982). Children's comprehension of "before" and "after" reinvestigated. *Journal of Child Language*, *9*, 381–402.

Tulving, E. (2005). Episodic memory and autonoesis: Uniquely human? In H. S. Terrace & J. Metcalfe (Eds.), *The missing link in cognition: Origins of self-reflective consciousness* (pp. 3–56). Oxford, UK: Oxford University Press.

Tulving, R. (2001). Episodic memory and common sense: How far apart? In A. Baddeley, M. Conway, & J. Aggleton (Eds.), *Episodic memory: New directions in research* (pp. 269–287). Oxford, UK: Oxford University Press.

Weist, R. M. (1986). Tense and aspect. In P. Fletcher & M. Garman (Eds.), *Language acquisition: Studies in first language development* (2nd ed., pp. 356–374). Cambridge, UK: Cambridge University Press.

Weist, R. M. (1989). Time concepts in language and thought: Filling the Piagetian void from two to five years. In I. Levin & D. Zakay (Eds.), *Time and human cognition: A life-span perspective* (pp. 63–118). Amsterdam: Elsevier.

Welch-Ross, M. (2001). Personalizing the temporally extended self: Evaluative self-awareness and the development of autobiographical memory. In C. Moore & K. Lemmon (Eds.), *The self in time: Developmental perspectives* (pp. 97–119). Mahwah, NJ: Erlbaum.

Wellman, H. M., Fabricius, W. V., & Sophian, C. (1985). The early development of planning. In H. M. Wellman (Ed.), *Children's searching: The development of search skills and spatial representation* (pp. 123–149). Hillsdale, NJ: Erlbaum.

Wellman, H. M., Somerville, S. C., & Haake, R. J. (1979). Development of search procedures in real-life spatial environments. *Developmental Psychology*, *15*, 530–542.

Welsh, M. C. (1991). Rule guided behaviour and self-monitoring on the Tower of Hanoi disk-transfer task. *Cognitive Development*, *6*, 59–76.

Zaitchik, D. (1990). When representations conflict with reality: The preschoolers' problem with false belief and false photographs. *Cognition*, *35*, 41–68.

Zelazo, P. D., & Somerville, J. A. (2001). Levels of consciousness of the self in time. In C. Moore & K. Lemmon (Eds.), *The self in time: Developmental perspectives* (pp. 229–252). Mahwah, NJ: Erlbaum.

Child Development, September/October 2008, Volume 79, Number 5, Pages 1498 – 1512

Effects of Family Violence on Psychopathology Symptoms in Children Previously Exposed to Maltreatment

Andrea Kohn Maikovich and Sara R. Jaffee
The University of Pennsylvania

Candice L. Odgers
University of California, Irvine

Robert Gallop
West Chester University

Although many studies suggest that family violence is associated with child psychopathology, multiple features of the home environment might account for this association, such as poverty and caregiver psychopathology. Studies are needed examining how change in psychopathology symptoms is affected by home violence, controlling for children's own developmental symptom histories and other predictors of psychopathology. This study used latent difference score structural equation modeling to test if witnessing home violence and/or experiencing harsh physical discipline predicted changes in psychopathology symptoms among 2,925 youth aged 5 – 16 years previously exposed to violence. Results demonstrated that harsh physical discipline predicted child-specific changes in externalizing symptoms, whereas witnessing violence predicted child-specific changes in internalizing symptoms across time. Implications for research and policy are discussed.

Many American children are exposed to violence in their homes. The U.S. Department of Health and Human Services, Administration for Children and Families, Children's Bureau (2005) reported that child protective service (CPS) agencies receive over 50,000 reports of suspected child maltreatment each week. After reviewing the available literature, Straus et al. (1998) concluded that approximately 5% of children are the victims of severe physical assault and that rates of physical abuse are potentially much higher depending on the exact definition of abuse utilized in any given study. These researchers also concluded that at least one in six American couples have engaged in partner-against-partner assault, and Straus (1992) estimated that 10 million American children witness domestic violence in their homes each year. Many children experience both forms of family violence (Slep & O'Leary, 2001). The majority of studies reviewed by Edleson (1999) found 30% – 60% co-occurrence rates, and several additional studies and reviews have also reported high rates of overlap (e.g., Appel & Holden, 1998; Moffitt & Caspi, 1998).

A plethora of studies now provide compelling evidence that children who are maltreated and children who witness family violence (e.g., adult partner violence) are at increased risk of psychopathology compared with children who are not exposed to violence (e.g., Jouriles, Murphy, & O'Leary, 1989; Kaplan, Pelcovitz, & Labruna, 1999; Kernic et al., 2003; Trickett & McBride-Chang, 1995; Wolfe, Crooks, Lee, McIntyre-Smith, & Jaffe, 2003). Jaffe, Wolfe, Wilson, and Zak (1986) found that abused boys and boys who had witnessed domestic violence had similar adjustment problem patterns and that both groups of boys had significantly more problems than a control group. In a recent meta-analysis, Wolfe et al. (2003) concluded that children who were both abused and exposed to domestic violence had higher levels of emotional and behavioral problems than did children who had only been exposed to domestic violence. Thus, children who are both victimized by violence and who witness domestic violence may be particularly vulnerable to developing psychopathology symptoms and are therefore an important population to study.

The extant literature suggests that there are several direct links between physical abuse and negative

---

This work was supported by grant HD050691 from the National Institute of Child Health and Human Development to Sara R. Jaffee. This document includes data from the National Survey of Child and Adolescent Well-Being, which was developed under contract with the Administration on Children, Youth, and Families, U.S. Department of Health and Human Services. The information and opinions expressed herein reflect solely those of the authors.

Correspondence concerning this article should be addressed to Andrea Kohn Maikovich, Department of Psychology, The University of Pennsylvania, 3720 Walnut Street, Philadelphia, PA 19104. Electronic mail may be sent to andreama@psych.upenn.edu.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7905-0018

child outcomes. For example, children who are physically abused are more likely to encode violence as an acceptable interaction strategy (Black & Newman, 1996) and have deficits in their attention self-regulatory abilities that cause them to overattend to anger cues (Shackman, Shackman, & Pollak, 2007). There are also indirect links between physical abuse and negative child outcomes. Abuse of a child is often associated with a host of suboptimal parenting practices that in and of themselves contribute to increases in children's internalizing and externalizing symptoms (e.g., Belsky, 1993; Burgess & Conger, 1978). For example, compared with nonabusive parents, abusive parents engage in significantly fewer positive interactions with their children, have interactions that more frequently contain angry and punitive statements, and interact less often with their children (Burgess & Conger, 1978; Vasta, 1982). Other researchers have found that nonabusive parents are more likely to use reasoning and simple commands and to choose disciplinary techniques that match the child's misbehavior, whereas abusive parents nearly always utilize punitive disciplinary techniques even for very minor child misdeeds (Trickett & Kuczynski, 1986).

Parenting practices and children's well-being might also be affected by the relatively high rates of stress and mental health problems that adult victims of family violence experience (Farver, Crooks, Lee, McIntyre-Smith, & Jaffe, 2005; Margolin & Gordis, 2000). For example, Farver et al. (2005) showed that the association between family and community violence and children's depressive symptoms was partially mediated by mothers' depressive symptoms. Furthermore, Wolfe, Crooks, Lee, McIntyre-Smith, and Jaffe (1985) found that the association between witnessing violence and children's problems was at least partially mediated by maternal stress and associated factors, and Holden and Ritchie (1991) found that maternal stress and paternal irritability were the two major predictors of behavior problems in children of battered women. Maternal depression may contribute to children's internalizing and externalizing symptoms by making mothers less consistently supportive and reliable caregivers (Katz & Gottman, 1997).

Although parenting problems might be one pathway by which family violence increases risk for child psychopathology, children who are exposed to violence are also likely to experience many other stressors that. are associated with elevated rates of psychopathology such as poverty, community violence, and poor schools (Hay & Jones, 1994; National Research Council, 1993). Even if families and children receive social services, children's homes frequently remain chaotic and violent, even if not sufficiently "abusive" to merit further action by CPS. The multiple stressors and risk factors for psychopathology that are associated with violence in the home need to be accurately measured and accounted for in research studies in order to isolate the effects of abuse and witnessing violence, as the association between family violence and children's problem behaviors may be accounted for by social disadvantage more generally.

Moreover, although family violence may lead to symptoms of depression, anxiety, substance use, or aggression in adult victims and perpetrators, adults who have histories of mental health problems and behavior problems are more likely to form relationships that become physically and psychologically abusive (Magdol, Moffitt, Caspi, & Silva, 1998). They are also more likely to abuse their children (Bland & Orn, 1986a, 1986b; Egami, Ford, Greenfield, & Crum, 1996). This implies that the effects of family violence on child well-being may be genetically mediated. Given that adult depression, anxiety, substance use, and antisocial behavior are all moderately heritable (McGuffin, Owen, & Gottesman, 2002), it is possible that family violence is merely a marker for genetic risk for psychopathology that parents transmit to children (DiLalla & Gottesman, 1991; Jaffee, Caspi, Moffitt, Taylor, & Arseneault, 2002, 2004).

In this study, we examined how the levels of violence children reported witnessing in their home and the amount of harsh physical discipline caregivers reported using with their children were associated with changes in the externalizing and internalizing symptoms of children previously exposed to violence. We hypothesized that both harsh physical discipline and witnessing violence would be associated with changes in externalizing and internalizing symptoms above children's expected trajectories in this sample of children previously exposed to violence. Externalizing and internalizing behavior problems typically decline from childhood to adolescence, although some children maintain consistently high levels of problem behaviors (Broidy et al., 2003; Chang, Halpern, & Kaufman, 2007). One goal of our study was to test whether reexposure to violence would account for individual differences in children's trajectories and explain why some children do not show the expected decline in problem behaviors.

Our study attempted to redress several methodological limitations of previous studies in this field. First, whereas many studies examine physical violence against children or adult domestic violence, these forms of violence tend to co-occur. Therefore, we estimate the unique effect of each type of violence on children's mental health. Second, whereas many studies estimate the concurrent or prospective associations between family violence and child psychopathology, our study examines the relationship between family violence and change over time in children's symptoms. This allows us to get a better handle on the direction of the relationship between family violence and children's problem behaviors and is a better test of whether family violence is a potential cause of children's psychopathology.

Method

Subjects

The National Survey of Child and Adolescent Well-Being (NSCAW) is a nationally representative sample of U.S. children who have had contact with CPS (Dowd et al., 2004). The full cohort includes 5,501 children (50% female), younger than 1 year to 16 years when first sampled, who were subjects of child abuse or neglect investigations conducted by CPS agencies from October 1999 to December 2000. The sample was selected using a two-stage stratified sample design. At the first stage, the United States was divided into nine sampling strata. Eight strata corresponded to the eight states with the largest child welfare caseloads, and the ninth stratum consisted of the remaining 38 states and the District of Columbia. Within each of the nine strata, primary sampling units (PSUs) were formed and randomly selected. PSUs were defined as geographic areas that encompassed the population served by a single CPS agency (e.g., counties). At the second stage, equal numbers of children were selected from each PSU, regardless of PSU size. Children were selected from eight mutually exclusive and exhaustive domains such that the final sample adequately represented relevant combinations of (a) infants versus children aged 1–14 years, (b) children receiving CPS-funded agency services versus children receiving no services, (c) children in out-of-home care versus children not in out-of-home care, and (d) children who were investigated for allegations of sexual abuse versus other forms of abuse or neglect. Additional information about the sample composition is available from Dowd et al. (2004).

Field staff completed 12 days of training on the protocol. At baseline (Wave 1), face-to-face interviews or assessments were conducted with children, their caregivers (e.g., biological parents, foster parents, custodial kin caregivers), their teachers (when children were of school age), and their caseworkers (when applicable). Follow-up interviews were conducted at 12, 18, and 36 months postbaseline. (Only the data from the first, third, and fourth waves of data collection were used in this article, as the data collection protocol was significantly different for the second wave.) Current caregivers were paid $50 for their participation and children were given gift certificates worth $10–$20.

The current analyses were restricted to children who were 5 years or older at Wave 1 (N = 2,925; 53% female) because younger children did not report on witnessing home violence. The racial/ethnic makeup of our sample was 46% White (non-Hispanic), 30% Black (non-Hispanic), 17% Hispanic, and 8% other races or ethnicities. Caseworkers reported on all subtypes of maltreatment children experienced and also designated the most serious type of maltreatment children experienced based on their review of the children's files and their ratings of the frequency, severity, and duration of each alleged or substantiated instance of abuse or neglect. For 40% of children, the most serious type of abuse was neglect; for 26%, it was physical; for 18%, it was sexual; for 9%, it was "other"; and for 8%, it was emotional. Within our sample, 34% of children experienced multiple types of maltreatment, and 62% of cases were substantiated. The mean child ages were 9.63 years (SD = 2.98) at Wave 1, 10.94 years (SD = 3.01) at Wave 3, and 12.24 years (SD = 2.99) at Wave 4.

Measures

Descriptive statistics are found in Table 1.

Externalizing symptoms. At each wave of data collection, caregivers were administered the Child Behavior Checklist (CBCL; Achenbach, 1991), which

Table 1
Descriptive Statistics for Variables Included in the Latent Difference Score Models

	M	*SD*	Range
Externalizing symptoms, baseline	16.13	11.52	0–60
Externalizing symptoms, 18 months	15.14	11.13	0–60
Externalizing symptoms, 36 months	14.59	11.14	0–65
Internalizing symptoms, baseline	10.40	8.67	0–58
Internalizing symptoms, 18 months	9.72	8.35	0–47
Internalizing symptoms, 36 months	9.48	8.09	0–54
Witnessed violence, baseline	3.57	2.72	0–12
Witnessed violence, 18 months	3.15	2.56	0–12
Witnessed violence, 36 months	2.80	2.55	0–12
Physical discipline, baseline	6.80	11.18	0–142
Physical discipline, 18 months	4.94	9.65	0–87
Physical discipline, 36 months	3.87	8.08	0–94
Income[a]	2.82[a]	1.41	1–5
Caregiver mental health	48.48	11.16	12–70
Child age, baseline	9.63	2.98	5–16

[a] Income was recoded such that 2 = $10,000–19,999 per year; 3 = $20,000–29,000 per year.

consists of 113 questions on a 3-point Likert-type scale (0 = *not true*, 1 = *somewhat or sometimes true*, 2 = *very true or often true*). We used the child's score on the CBCL Externalizing scale (which taps delinquent and aggressive behavior) as our measure of externalizing symptoms. Internal consistency reliability was high for this scale (α = .92). At each wave of analysis, 33%, 30%, and 28% of children had clinically significant externalizing scores, respectively (i.e., *t* scores at or above 65; Achenbach, 1991).

Internalizing symptoms. We used children's scores on the Internalizing scale of the CBCL (which comprises withdrawn behavior, somatic complaints, and anxious/depressed domains) as our measure of children's internalizing symptoms. Internal consistency reliability was high for this scale (α = .90). At each wave of analysis, 25%, 18%, and 17% of children had clinically significant internalizing scores, respectively (i.e., *t* scores at or above 65; Achenbach, 1991).

Witnessed violence. Children were administered the Violence Exposure Scale for Children (Fox & Leavitt, 1995), during which they were asked 23 questions regarding their exposure to violence and criminal events. Children younger than 11 years were also shown a cartoon depiction of each act, and all children were asked how often they had witnessed each act *at home* (never, once, a few times, or lots of times). They were then asked if they had witnessed the act in the past month (1 = *yes*, 2 = *no*). Internal consistency on this measure ranged from αs = .72 to .86 in a sample of inner-city minority preschool children (Shahinfar, Fox, & Leavitt, 2000).

In two-factor analyses, Raviv et al. (2001) and Raviv, Shimoni, Fox, & Leavitt (1999) identified a "mild" and a "severe" violence categorization scheme for the items on this measure. There were six types of violent acts that loaded onto the *witnessing mild violence* category. These included observing an adult yell at someone, observing an adult throw something at another person, watching an adult push or shove someone, watching an adult slap someone, observing an adult beat someone up, and observing another child getting spanked. The six items that loaded onto the *witnessing severe violence* category included observing a person steal things from another person, seeing an adult point a knife or gun at someone, observing someone stab another person, seeing someone shoot another person with a gun, observing someone getting arrested, and seeing someone deal drugs. We summed the number of different types of incidents children reported witnessing in the past month to create a total witnessed violence scale. Thus, children received a score between 0 and 12. In the NSCAW sample, internal consistency was high for the total score (α = .96). Across waves of data collection, between 71% and 80% of children did not witness any of the severe violence category events.

Caregiver use of harsh physical discipline. Caregivers were administered the Parent–Child Conflict Tactics Scales (CTSPC; Straus et al., 1998), which ask caregivers how often they used 22 disciplinary practices in the past year (0 = *never or not in the past 12 months*; 1 = *1 time*; 2 = *2 times*; 4 = *3–5 times*; 8 = *6–10 times*; 15 = *11–20 times*; 25 = *more than 20 times*). We used the caregivers' total score on the physical assault scale as our measure of harsh physical discipline, after dividing the total score by 10 in order to make the scale commensurate with that of our other variables. Examples of items on this scale range from spanking children with a bare hand to choking children. In the NSCAW sample, internal consistency for the measure was α = 0.92. Across waves of data collection, 86%–91% of children did not experience any of the "severe" or "very severe" physical discipline category events, so scores on this measure primarily reflected corporal punishment as opposed to physical abuse.

Income. Families were classified into five categories based on total family income in 1 year (1 = $0–$9,999; 2 = $10,000–$19,999; 3 = $20,000–$29,000; 4 = $30,000–$39,999; 5 = $40,000 and over).

Caregiver's mental health. Caregivers were administered the short-form health survey (Ware, Kosinski, & Keller, 1998), which assesses mental and physical health. We used the mental health subscale as our measure of caregiver mental health. This subscale has been shown to differentiate groups known to differ in terms of the presence and seriousness of their mental health problems (Ware, Kosinski, & Keller, 1996). Some of the questions on this subscale asked caregivers how often they felt a certain way, such as "downhearted or blue," and others asked how much emotional problems interfered with their daily life and social functioning (1 = *all the time*; 2 = *most of the time*; 3 = *a good bit of the time*; 4 = *some of the time*; 5 = *a little of the time*; 6 = *none of the time*). Caregivers were asked to respond to all questions based on their experiences during the past month. Age- and gender-standardized scores were created, with higher scores indicating better mental health. Internal consistency for the mental health scale in the NSCAW sample was acceptable (α = .79).

Results

Statistical Analyses

In this study, we used a series of latent difference score (LDS) models (Hamagami & McArdle, 2001;

King et al., 2006) to examine our hypothesis that witnessing violence in the home and being subjected to harsh physical discipline would predict changes in the externalizing and internalizing symptoms of children previously exposed to violence. LDS modeling has recently been put forward as an innovative methodological approach for advancing the longitudinal study of trauma (King et al., 2006). LDS models are part of a larger family of longitudinal structural models designed to assess growth and change and offer an advantage over traditional change score approaches by partitioning the true score from the error variance in a time series and producing an optimally reliable latent change score over assessment points. In the present study, LDS models provide a dynamic means of testing whether violence exposure predicts changes in children's psychopathology symptoms across time. The structural equation modeling (SEM) framework adopted here also offers the advantage of accounting for the nonindependence of children's violence exposure over time, and allows us to address the question of whether violence exposure predicts *changes* in children's psychopathology symptoms, after accounting for children's prior symptoms and other key predictors of psychopathology.

The analyses proceeded in three steps. The same steps were followed for both externalizing and internalizing symptoms. First, a baseline LDS model was fit to the data to derive the latent symptom change scores. Second, we built on the baseline models to investigate whether witnessing violence in the home and/or experiencing harsh physical discipline predicted changes in children's externalizing and internalizing symptoms relative to their own developmental histories after accounting for other important child and family risk factors for psychopathology. Third, we ran multiple group models to examine whether children's previous exposure to physical abuse moderated the relationship between violence exposure and changes in psychopathology symptoms.

Models were estimated using Amos version 5.0 (Arbuckle, 2003). In order to evaluate model fit, we used the model chi-square test; however, because the interactive effect of sample size and model error on this test typically causes the model chi-square to be statistically significant with large samples even when the model represents a close fit to the data (MacCallum, 1990), we employed three additional fit indices: the comparative fit index (CFI; Bentler, 1990), the Tucker–Lewis index (TLI; Tucker & Lewis, 1973), and the root mean square error of approximation (RMSEA; Steiger, 1990). Model chi-square values with accompanying p values greater than .05 indicate a good model fit. CFI and TLI values greater than 0.95 and RMSEA values less than 0.05 indicate a good fit; CFI and TLI values between 0.90 and 0.95 and RMSEA values between 0.05 and 0.08 indicate an acceptable fit. (For a discussion of the various fit indices, see Browne & Cudeck, 1992; Medsker, Williams, & Holahan, 1994).

Missing data. Missing data ranged from 0% to 36% across variables (*Mdn* = 16%). The most frequently missing data were for harsh physical discipline (26% missing at baseline, 36% missing 18 months postbaseline, and 34% missing 36 months postbaseline). Missing data were addressed using full information maximum likelihood estimation under the assumption that the data were missing completely at random or for reasons that could be explained by other variables included in the model (Little & Rubin, 1987). In technical terms, a covariance coverage matrix is created that provides the proportion of available observations for each time point and pairs of time points. This method is a widely accepted method of addressing missing data within an SEM framework while allowing for the inclusion of all available data points (Arbuckle, 1996; Enders, 2001; Raykov, 2005), and in the present study retained all 2,925 participants for the analyses.

Key Variable Correlations

The correlations between all variables in the models are displayed in Table 2. Within-wave correlations between witnessing violence and children's internalizing and externalizing symptoms were consistently small, but significant (*r*s = .09–.17). Within-wave correlations between harsh physical discipline and children's internalizing and externalizing symptoms were small to moderate in magnitude (*r*s = .11–.25). Within-wave correlations between witnessing violence and harsh physical discipline were small, but statistically significant (*r*s = .07–.10)

Externalizing Symptoms

Baseline model. We estimated a baseline LDS model for the repeated measure of externalizing symptoms as assessed at baseline, 18 months postbaseline, and 36 months postbaseline, thereby creating two LDS that represent the change in externalizing symptoms between data collection waves (Figure 1). Alpha was set at .05, and thus all results reported as significant are $p < .05$ or better. The baseline model fit the observed data well, and the model fit was significant, $\chi^2(1) = 23.31$, CFI = 0.99, TLI = 0.95, RMSEA = 0.09. The model-implied mean level of externalizing symptoms decreased significantly from 16.15 to 14.43

Table 2

Correlation Matrix of All Variables in Latent Difference Score Models

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. Gender	1.00															
2. Age	.07	1.00														
3. Income, W1	−.01	.05	1.00													
4. CGMH, W1	−.01	−.03	.17	1.00												
5. Ph. disc., W1	−.09	−.18	−.07	−.14	1.00											
6. Ph. disc., W3	−.05	−.18	NA	NA	.50	1.00										
7. Ph. disc., W4	−.06	−.21	NA	NA	.43	.54	1.00									
8. Wit. viol., W1	.02	−.08	−.01	−.05	**.10**	.08	.07	1.00								
9. Wit viol., W3	.05	−.03	NA	NA	.08	**.07**	.07	.32	1.00							
10. Wit. viol., W4	.08	.04	NA	NA	.04	.03	**.08**	.26	.43	1.00						
11. Int., W1	.01	.16	.01	−.31	**.11**	.06	.02	**.10**	.09	.11	1.00					
12. Int., W 3	.01	.13	NA	NA	.07	**.14**	.10	.10	**.14**	.13	.66	1.00				
13. Int., W4	.03	.10	NA	NA	.06	.10	**.16**	.08	.11	**.16**	.46	.57	1.00			
14. Ext., W1	−.13	.13	0	−.26	**.25**	.17	.13	**.09**	.12	.12	.65	.43	.37	1.00		
15. Ext., W3	−.11	.11	NA	NA	.16	**.23**	.18	.07	**.15**	.13	.37	.66	.45	.60	1.00	
16. Ext., W4	−.11	.08	NA	NA	.16	.19	**.25**	.07	.11	**.17**	.32	.41	.67	.52	.65	1.00

Note. The correlations in bold are the correlations between the main outcome variables and the main predictor variables of interest for the research question. *rs* ≥ .04 are significant at the $p < .05$ level; *rs* ≥ .05 are significant at the $p < .01$ level. W1 = Wave 1; W3 = Wave 3; W4 = Wave 4; CGMH = caregiver mental health; Ph. disc. = physical discipline; Wit. viol. = witnessed violence; Int. = internalizing symptoms; Ext. = externalizing symptoms; NA = not applicable.

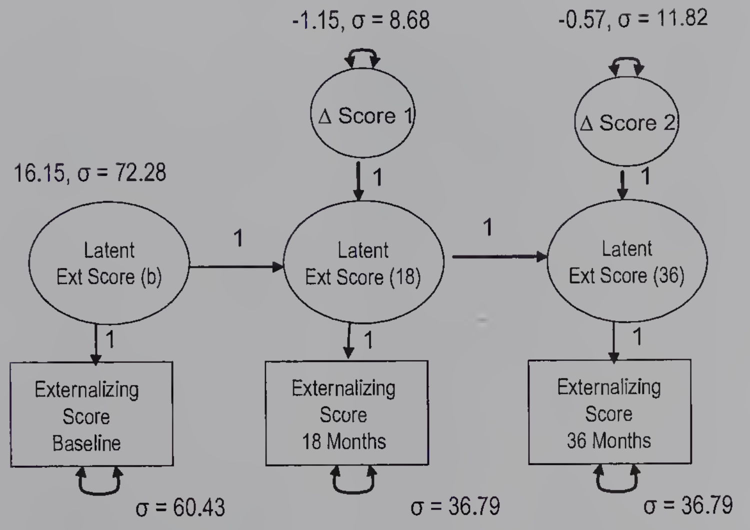


Figure 1. Baseline latent difference score (LDS) model for externalizing symptoms.
Note. In this model, Δ Score 1 is the LDS representing the change in externalizing symptoms from baseline to 18 months. Δ Score 2 is the LDS representing the change in externalizing symptoms from 18 months to 36 months. Double-headed arrows above the LDS represent the variance (σ) of the difference scores. Observed scores (externalizing scores) reflect true scores plus error. Double-headed arrows attached to the observed scores represent the error variance (σ) of the externalizing scores at each assessment.

Table 3
Latent Difference Score Model for Witnessed Violence and Physical Discipline on Externalizing Symptoms

Parameter estimates	Estimates	CR
Baseline model		
Externalizing symptoms intercept	16.15	75.48***
Initial status variance	72.28	27.58***
Difference score$_1$ mean	−1.15	−5.65***
Difference score$_1$ variance	8.68	4.09***
Difference score$_2$ mean	−0.58	−2.99***
Difference score$_2$ variance	11.82	3.87***
Model with covariates		
Exposure to violence		
Witness$_1$ → Diff$_1$	0.04	1.02
Witness$_3$ → Diff$_2$	0.03	1.02
Physical discipline$_1$ → Diff$_1$	0.12	2.95**
Physical discipline$_3$ → Diff$_2$	0.12	2.95**
Witness$_1$ → Witness$_3$	0.38	25.79***
Witness$_3$ → Witness$_4$	0.36	25.79***
Physical discipline$_1$ → Physical discipline$_3$	0.55	37.68***
Physical discipline$_1$ → Physical discipline$_3$	0.47	37.68***
Covariates		
Age → Externalizing intercept	0.15	7.10***
Gender → Externalizing intercept	−0.16	−7.79***
Caregiver mental health → Externalizing intercept	−0.24	−10.97***
Income → Externalizing intercept	0.03	1.32

Fit indices	Baseline model	Model with covariates
χ^2/df	23.31/1	315.60/26
CFI	0.99	0.93
TLI	0.95	0.88
RMSEA	0.09	0.07

Note. All estimates are standardized estimates. CR is critical ratio (estimate/standard error); values > 1.96 are statistically significant at the $p < .05$ level. Diff = difference score; CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = root mean square error of approximation. $^{**}p < .01$. $^{***}p < .001$.

externalizing symptoms from baseline to 36 months postbaseline. As displayed in Table 3, results from the model indicated that there was enough variability in how children changed between assessment points (i.e., variance in the difference scores) to justify testing whether the variation in children's latent externalizing change scores could be explained by witnessing violence and/or experiencing harsh physical discipline.

Does exposure to violence predict changes in children's externalizing scores? Next, we built on the baseline LDS model (Figure 2) to examine the effects of witnessing home violence and experiencing harsh physical discipline on children's externalizing symptoms, after controlling for child gender, family income, caregiver psychopathology, and child age (for ease of presentation, these covariates do not appear in the figure, although the parameter estimates appear in Table 3). The model fit the data adequately, and the model fit was significant, $\chi^2(32) = 315.60$, CFI = 0.93, TLI = 0.88, RMSEA = 0.07. Although the TLI has fallen below 0.90, the CFI and RMSEA were still good for a model of this complexity. There was no relationship between witnessing violence and change in externalizing symptoms across assessments ($\beta_1 = 0.04$, $z = 1.02$, *ns*; $\beta_2 = 0.03$, $z = 1.02$, *ns*). However, experiencing harsh physical discipline did predict significant changes in children's externalizing symptoms across assessments ($\beta_1 = 0.12$, $z = 2.95$; $\beta_2 = 0.09$, $z = 2.95$). Although, on average, externalizing problems declined over time, children who experienced relatively high levels of harsh physical discipline showed more gradual declines and, in some cases, increases in externalizing problems from baseline to 36 months postbaseline. This effect remained statistically significant after controlling for child and family risk factors.

We also ran a multiple group model in order to examine whether maltreatment subtype moderated the relationship between exposure to violence (harsh physical discipline or witnessing violence) and changes in externalizing problems. Specifically, we were interested in whether the effects of exposure to violence differed for children who had been physically abused versus children who had experienced

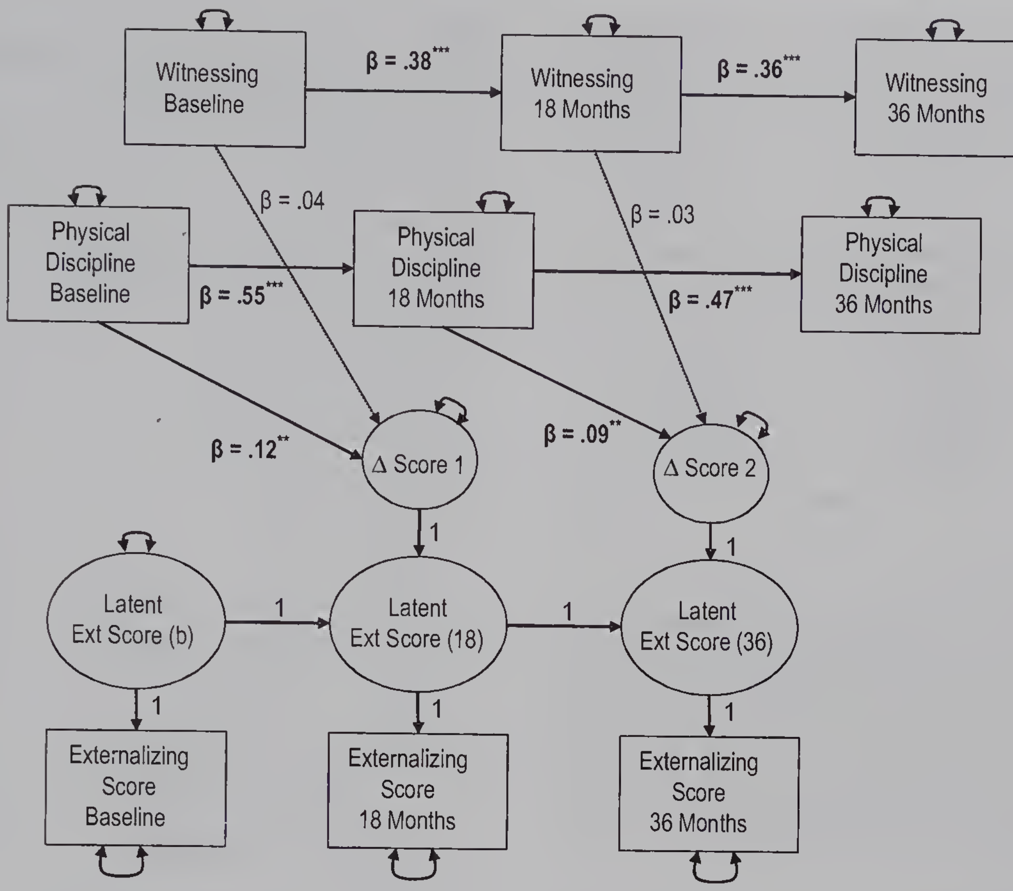

Witnessing Baseline
β = .38***
Witnessing 18 Months
β = .36***
Witnessing 36 Months
β = .04
β = .03
Physical Discipline Baseline
β = .55***
Physical Discipline 18 Months
β = .47***
Physical Discipline 36 Months
β = .12**
Δ Score 1
β = .09**
Δ Score 2
1
Latent Ext Score (b)
1
Latent Ext Score (18)
1
Latent Ext Score (36)
1
1
1
1
Externalizing Score Baseline
Externalizing Score 18 Months
Externalizing Score 36 Months

Figure 2. Latent difference score (LDS) model for the relationship between witnessing violence, physical discipline, and externalizing symptoms (*N*= 2,925).

Note. In this model, Δ Score 1 is the LDS representing the change in externalizing symptoms from baseline to 18 months. Δ Score 2 is the LDS representing the change in externalizing symptoms from 18 months to 36 months. The nonindependence of witnessing violence across assessments is represented by the regression of witnessing violence on itself across baseline, 18-month and 36-month assessments. The nonindependence of experiencing physical discipline across assessments is represented by the regression of physical discipline on itself across baseline, 18-month and 36-month assessments. Double-headed arrows represent the variance of latent and observed variables. Although not shown in the model for sake of clarity, the correlation between witnessed violence and harsh physical discipline was small but significant ($r = .06, p < .05$). Parameter estimates in Figure 2 are adjusted for age, gender, caregiver mental health, and income. Parameter estimates for the covariate controls are not included in Figure 2 but are available in Table 3.

$^{**}p < .01$. $^{***}p < .001$.

other forms of abuse. We hypothesized that exposure to violence might evoke a response that either reflected habituation or sensitization processes in youth who had already experienced physical violence. However, the model in which the effects of exposure to violence on children's externalizing symptoms were constrained to be the same for children who were physically abused versus children who were not physically abused was not a significantly worse fit than the model in which the effect of exposure to violence was free to vary across abuse subgroups, $\chi^2(2) = 0.7$, *ns*. Full analyses are available from the authors upon request.

Internalizing Symptoms

Baseline model. We estimated a baseline LDS model for the repeated measure of internalizing symptoms as assessed at baseline, 18 months postbaseline, and 36 months postbaseline, thereby creating two LDS that represent the change in internalizing symptoms between data collection waves (Figure 3). The baseline model fit the observed data adequately and the model fit was significant, $\chi^2(1) = 32.53$, CFI = 0.98, TLI = 0.90, RMSEA = 0.10. The model-implied mean level of internalizing symptoms decreased significantly from 10.41 to 9.32, internalizing symptoms from baseline to 36 months postbaseline. As displayed in Table 4, results from this baseline model indicated that there was enough variability in how children changed between assessment points to justify testing whether the variation in children's latent internalizing change scores could be explained by witnessing violence and/or experiencing harsh physical discipline.

Does exposure to violence predict changes in children's internalizing scores? Next, we built on the baseline LDS model (Figure 4) to examine the effects of witnessing home violence and experiencing harsh physical discipline on children's internalizing

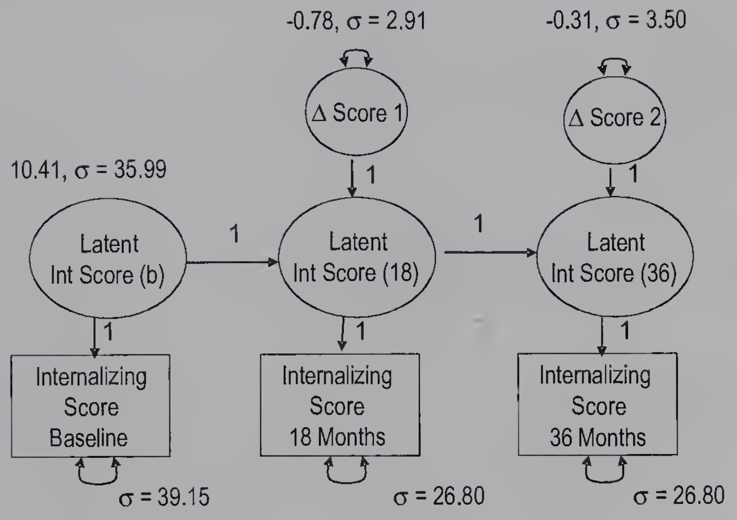


Figure 3. Baseline latent difference score (LDS) model for internalizing symptoms.
Note. In this model, Δ Score 1 is the LDS representing the change in internalizing symptoms from baseline to 18 months. Δ Score 2 is the LDS representing the change in internalizing symptoms from 18 months to 36 months. Double-headed arrows above the LDS represent the variance (σ) of the difference scores. Observed scores (internalizing scores) reflect true scores plus error. Double-headed arrows attached to the observed scores represent the error variance (σ) of the internalizing scores at each assessment.

Table 4
Latent Difference Score Model for Witnessed Violence and Physical Discipline on Internalizing Symptoms

Parameter estimates	Estimates	CR
Baseline model		
Internalizing symptoms intercept	10.41	64.64***
Initial status variance	35.99	25.69***
Difference score$_1$ mean	−0.78	−4.74***
Difference score$_1$ variance	2.91	2.35*
Difference score$_2$ mean	−0.31	−1.95***
Difference score$_2$ variance	3.50	1.86
Model with covariates		
Exposure to violence		
Witness$_1$ → Diff$_1$	0.11	2.91**
Witness$_3$ → Diff$_2$	0.12	2.91**
Physical discipline$_1$ → Diff$_1$	0.07	1.56
Physical discipline$_3$ → Diff$_2$	0.06	1.56
Witness$_1$ → Witness$_3$	0.38	25.81***
Witness$_3$ → Witness$_4$	0.36	25.81***
Physical discipline$_1$ → Physical discipline$_3$	0.55	37.64***
Physical discipline$_1$ → Physical discipline$_3$	0.47	37.64***
Covariates		
Age → Internalizing intercept	0.19	9.00***
Gender → Internalizing intercept	−0.01	−.29
Caregiver mental health → Internalizing intercept	−0.31	−13.94***
Income → Internalizing intercept	0.06	2.46*

Fit indices	Baseline model	Model with covariates
χ^2/df	32.53/1	361.60/26
CFI	0.98	0.92
TLI	0.90	0.85
RMSEA	0.10	0.07

Note. All estimates are standardized estimates. CR is critical ratio (estimate/standard error); values > 1.96 are statistically significant at the $p < .05$ level. Diff = difference score; CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = root mean square error of approximation. *$p < .05$. **$p < .01$. ***$p < .001$.

symptoms, after controlling for child gender, family income, caregiver psychopathology, and child age (for ease of presentation, these covariates do not appear in the figure, although the parameter estimates appear in Table 4). The model fit the data adequately, and the model fit was significant, $\chi^2(32) = 361.60$, CFI = 0.92, TLI = 0.85, RMSEA = 0.07. Although the TLI has fallen below .90, the CFI and RMSEA were still good for a model of this complexity. There was no relationship between harsh physical discipline and change in internalizing symptoms across assessments ($\beta_1 = 0.07$, $z = 1.56$, *ns*; $\beta_2 = 0.06$, $z = 1.56$, *ns*). However, witnessing violence in the home did predict significant changes in children's internalizing symptoms across assessments ($\beta_1 = 0.11$, $z = 2.91$; $\beta_2 = 0.12$, $z = 2.91$). Although, on average, internalizing problems declined over time, children who witnessed relatively high levels of home violence showed more gradual declines and, in some cases, increases in internalizing problems from baseline to 36 months postbaseline. This effect remained statistically significant after controlling for child and family risk factors.

We also ran a multiple group model in order to examine whether maltreatment subtype moderated the relationship between exposure to violence and internalizing problems. Again, we were specifically interested in whether the effects of exposure to violence differed for children who had been physically abused versus children who had not been physically abused. The model in which the effect of exposure to violence on children's internalizing symptoms was constrained to be the same for children who were physically abused versus children who were not physically abused was not a significantly worse fit than the model in which the effect of exposure to violence was free to vary across abuse subgroups, $\chi^2(2) = 3.30$, *ns*. Full analyses are available from the authors upon request.

Discussion

Exposure to violence was associated with changes in children's psychopathology symptoms in this study of children previously reported to CPS. Specifically, harsh physical discipline was associated with

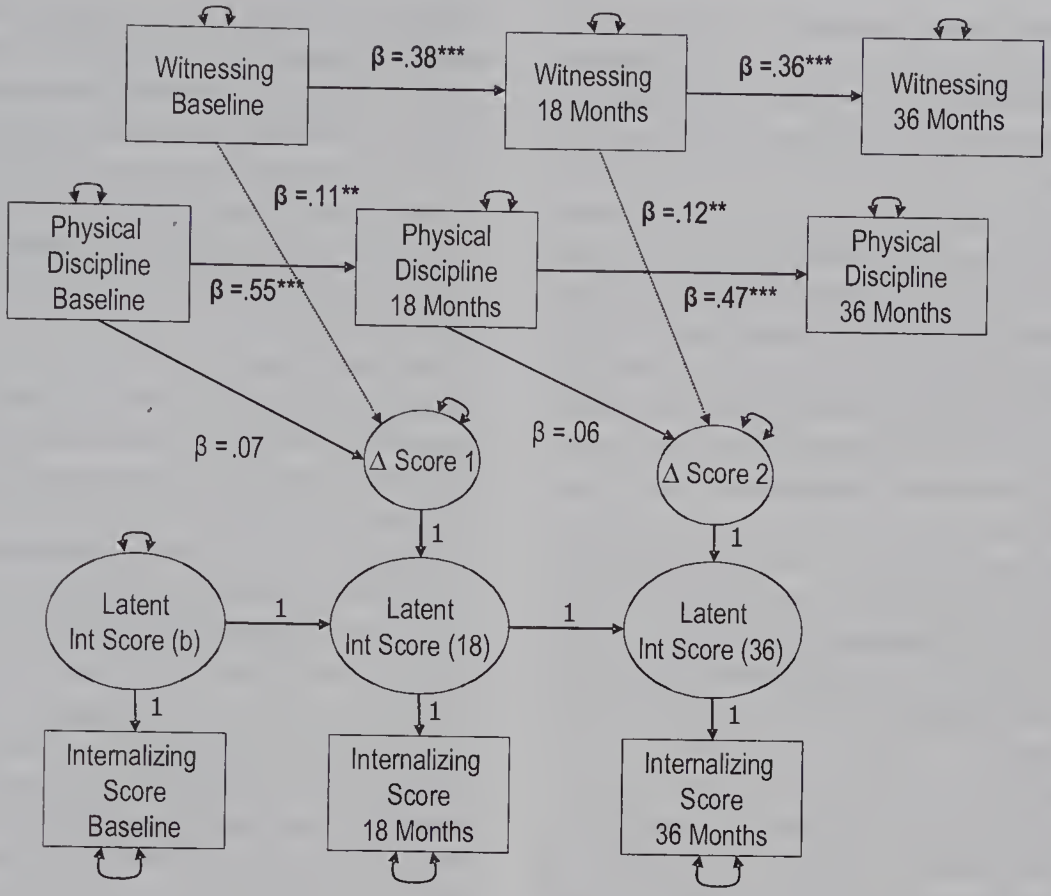


Figure 4. Latent difference score (LDS) model for the relationship between witnessing violence, physical discipline, and internalizing symptoms ($N = 2{,}925$).

Note. In this model, Δ Score 1 is the LDS representing the change in internalizing symptoms from baseline to 18 months. Δ Score 2 is the LDS representing the change in internalizing symptoms from 18 months to 36 months. The nonindependence of witnessing violence across assessments is represented by the regression of witnessing violence on itself across baseline, 18-month and 36-month assessments. The nonindependence of experiencing physical discipline across assessments is represented by the regression of physical discipline on itself across baseline, 18-month and 36-month assessments. Double-headed arrows represent the variance of latent and observed variables. Although not shown in the model for sake of clarity, the correlation between witnessed violence and harsh physical discipline was small but significant ($r = .06$, $p < .05$). Parameter estimates in Figure 2 are adjusted for age, gender, caregiver mental health, and income. Parameter estimates for the covariate controls are not included in Figure 4 but are available in Table 4.
$^{**}p < .01$. $^{***}p < .001$.

deviations from the normative sample trajectory of declining externalizing symptoms and witnessing home violence was associated with deviations from the normative sample trajectory of declining internalizing symptoms. These effects of family violence were significant even controlling for two potential confounds—family income and caregiver mental health—and after controlling for the child's gender and age.

This study had several methodological strengths. First, the sample size was large ($N = 2{,}925$) and nationally representative of children involved with CPS in the United States. Second, whereas current caregivers reported on their children's externalizing and internalizing symptoms, the children themselves reported on the levels of violence they witnessed in the home. Past studies suggest that caregivers significantly underestimate the amount of domestic violence to which their children are exposed (Jaffe, Wolfe, & Wilson, 1990; O'Brien, John, Margolin, & Erel, 1997). Third, we utilized an SEM approach (LDS modeling) that allowed us to examine the association between witnessing violence, harsh physical discipline, and child psychopathology after controlling for normative developmental changes in children's externalizing and internalizing symptoms over a 36-month period as well as a number of potential confounds. The modeling technique also allowed us to account for the nonindependence of exposure to violence over time. Many past studies in this field have lacked either longitudinal data or sufficient power to take into account children's normative symptom changes and to adequately control for important confounds. Furthermore, because violent discipline tactics and domestic violence witnessed by children frequently co-occur, it is important to attempt to tease apart the effects of each type of violence so as to inform future studies and interventions. Our modeling technique allowed us to examine these questions.

Our findings were largely consistent with the extant literature showing that violence exposure is associated with psychopathology symptoms in children. Additionally, we ruled out three alternative explanations for why harsh physical discipline and witnessing violence were associated with child psychopathology.

First, in light of the extensive body of literature showing that mental health problems run in families (e.g., Kendall, 2000), it is not surprising that in our sample, caregivers with poor mental health were raising children who had elevated levels of externalizing and internalizing symptoms. However, exposure to violence was still predictive of children's psychopathology symptom changes even after controlling for caregiver mental health. Thus, we can conclude that our results are not simply reflective of a spurious association between family violence and child psychopathology generated by caregiver mental health problems or by transmission of a genetic vulnerability for mental health problems from parents to children.

Second, multiple studies have shown that poverty is associated with elevated rates of child psychopathology (Bradley & Corwyn, 2002; Brooks-Gunn & Duncan, 1997; Costello et al., 2003; McLoyd, 1997). In addition, studies suggest that low-socioeconomic status (SES) parents are more likely than high-SES parents to engage in harsh or neglectful parenting styles, which are associated with more childhood problem behaviors (e.g., Bradley & Corwyn, 2002). However, harsh physical discipline was uniquely associated with externalizing symptoms and witnessing violence was uniquely associated with internalizing symptoms even controlling for family income.

Third, it is conceivable that changes in children's symptoms merely reflect the passage of time and normative trajectories. However, in our models, we controlled for children's expected symptom trajectories and showed that exposure to violence predicted deviations from these trajectories.

Implications for Research and Theory

We suggest that more research is needed to better understand why family violence increases risk for children's mental health problems. For example, although NSCAW children reported on the violent events they witnessed, they were not asked how they encoded or interpreted the events (or indeed how they encoded physical punishment). Upon reviewing the literature on children who witness domestic violence, Black and Newman (1996) concluded that researchers should give more consideration to the ways in which children cognitively interpret the violent events they witness, and Crockenburg and Forgays (1996) found that children's negative emotional reactions to their fathers during marital arguments independently predicted children's behavioral adjustment. Future studies might look more closely at how children who have previously been exposed to violence encode violent events they witness and the ways in which their parents discipline them in order to help better understand the long-term effects of family violence.

Children who have been exposed to violence severe enough to bring them to the attention of CPS, such as the children in this sample, might be particularly vulnerable to encoding the incidents they witness or experience themselves in a way that contributes to an internal representation of the world as extremely violent, unsafe, and unpredictable and, consequently, might have stronger and more negative emotional reactions to violence and conflict. Several studies have shown that children's histories, including their past exposure to parental conflict, influence how and how much witnessing marital conflict affects them (e.g., Cummings, Vogel, Cummings, & El-Sheikh, 1989; Davies, Myers, Cummings, & Heindel, 1999). Specifically, children who have been exposed to marital conflict more frequently in the past have more negative emotional reactions to new incidences of interparental conflicts (Davies & Cummings, 1998). If children who have already been victims of maltreatment encode adult domestic violence as especially threatening, this may increase their vulnerability to internalizing and externalizing problems. Pollak's work with physically abused children, for example, suggests that abused children display increased anticipatory monitoring in response to angry interpersonal situations in the environment and display a deficit in their ability to regulate their arousal (Pollak, Vardi, Bechner, & Curtin, 2005).

There are two aspects of the results that should be examined further in future studies. First, physical abuse did not emerge as a significant moderator of the relationship between family violence and children's internalizing and externalizing problems in these analyses. However, the relationship between family violence and other types of psychopathology (or specific symptoms of internalizing and externalizing problems) may indeed depend on the type of abuse the child experienced. Second, experiencing harsh physical discipline was associated with changes in externalizing symptoms, whereas witnessing violence was associated with changes in internalizing symptoms. As reviewed above, other studies have demonstrated that both forms of family violence are associated with both internalizing and externalizing problems. Thus, our findings will need to be replicated in other samples before we can conclude with certainty that different types of family violence have specific effects on different types of child problem behaviors. To the extent that our finding is real, however, one possibility is that the different cognitive meanings these two forms of violence have for children lead to different manifestations of distress.

Finally, other research conducted with the NSCAW sample (Jaffee & Gallop, 2007) and with other samples (Cicchetti & Rogosch, 1997) has found that formerly abused children who are "free" of significant mental health problems at one point in time rarely maintain good mental health over time. Our study suggests that one reason may be that children with histories of abuse are often reexposed to various forms of violence that predict the recurrence of mental health problems. Again, more research is needed in this area.

Implications for Practice and Policy

Victims of abuse and neglect are at elevated risk of witnessing and experiencing other forms of family violence. The results of this study suggest that intervention efforts to reduce rates of mental health problems in child victims of maltreatment must focus not only on protecting children from revictimization but must also work to decrease even nonabusive forms of physical discipline and the amount of adult domestic violence children witness in their homes. Intervention efforts might help parents manage relationship conflict or, at the very least, educate parents about the importance of buffering children from exposure to conflict.

Past research on children's emotional responses to witnessing marital conflict suggests that children are less distressed by nonviolent conflict when conflicts are resolved with a compromise or an apology (Cummings, Ballard, El-Sheikh, & Lake, 1991). Future studies could explore whether this finding holds for children who witness *violent* conflict in the home. The impact of various resolution strategies might differ for children who witness violent forms of conflict, in that children who witness a continuous cycle of domestic violence and resolution of violence (such as an abusive father who breaks down into tears and apologizes to the child's mother but then abuses her again the next week) might actually normalize violence and be more prone to future problems. Given past research suggesting that children who witness domestic violence are at risk of encoding violence as an acceptable behavior in relationships (Black & Newman, 1996), resolution-improving interventions might emphasize not only that the behavior was unacceptable but also model for children more appropriate conflict-resolution strategies and disciplinary techniques.

In sum, our findings suggest that reducing the amount of violence that children witness in the home and the frequency with which their parents use harsh physical discipline could be important foci of effective family therapy interventions and prevention programs for this vulnerable population. Preventing abuse from reoccurring may not be enough to divert children from the maladaptive trajectories on which their past abuse and home environments contributed to placing them in the first place.

Limitations

First, although harsh physical discipline and witnessing violence were significant and unique predictors of psychopathology symptoms, the zero-order correlations among these variables were modest. Second, although reliability and validity statistics for the short-form health survey (from which we created our caregiver mental health variable) were good, the measure is a broad and general measure of mental health. It does not ask questions about specific mental illnesses but rather asks more about daily impairment due to emotional problems. Third, it is possible that some parents underreported their use of violent disciplinary techniques or children's problems, given that they had already been brought to the attention of CPS. Although efforts were made to minimize this possibility by administering questions about physical discipline via an audio computer-assisted interview and interviewers were thus unaware of parents' responses, Tourangeau and Yan's (2007) meta-analysis found that computer administration did not lead to significantly increased reporting of sensitive information by research participants and, indeed, rates of missing data were higher for this variable than for others.

Fourth, although we included a number of variables in the LDS models that might have accounted for the association between children's experiences of family violence and their problem behaviors, it is possible that important confounds were not assessed (e.g., measures of neighborhood violence). Moreover, although we attempted to rule out the possibility that poverty accounted for the association between family violence and child psychopathology, it is possible that there was insufficient variability in income in this high-risk sample to adequately test this hypothesis. Fifth, although youth reported on witnessed violence in the home, caregivers reported on children's experiences of harsh physical discipline as well as child psychopathology symptoms, introducing shared informant variance. Sixth, although our models tested whether higher levels of violence predicted changes in psychopathology symptoms, we did not explicitly test whether changes in harsh discipline and witnessed violence predicted changes in psychopathology. Future research should examine this question, especially research intervention designs that alter parenting practices.

Children continue to be victimized by and exposed to domestic violence at alarming rates. In all likelihood, these children are at increased risk of psychopathology symptoms not only because they have been maltreated or exposed to violence but also because of their family's poverty status, the neighborhood violence they witness, the poor schools they attend, and the inadequacy of their housing. Thus, studies that explore how and why specific aspects of these early childhood environments affect children's mental health are important foundational work that will allow for the development of more efficacious, cost-effective, and targeted interventions. Studies that test alternative hypotheses about links between family violence and child psychopathology help researchers develop more specific hypotheses about mechanisms that link early childhood stressors to child socioemotional development.

References

Achenbach, T. M. (1991). *Manual for the Child Behavior Checklist/4-18 and 1991 Profile*. Burlington: Department of Psychiatry, University of Vermont.

Appel, A. E., & Holden, G. W. (1998). The co-occurrence of spouse and physical child abuse: A review and appraisal. *Journal of Family Psychology, 12*, 578–599.

Arbuckle, J. L. (1996). Full information estimation in presence of incomplete data. In G.A. Marcoulides & R.E. Schumacker (Eds.), *Advanced structural equation modeling, issues and techniques* (pp. 243–277). Mahwah, NJ: Erlbaum.

Arbuckle, J. L. (2003). *Amos 5.0 update to the Amos user's guide*. Chicago: SmallWaters.

Belsky, J. (1993). Etiology of child maltreatment: A developmental-ecological analysis. *Psychological Bulletin, 114*, 413–434.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238–246.

Black, D., & Newman, M. (1996). Children and domestic violence: A review. *Clinical Child Psychology and Psychiatry, 1*, 79–88.

Bland, R., & Orn, H. (1986a). Family violence and psychiatric disorder. *Canadian Journal of Psychiatry, 31*, 129–137.

Bland, R., & Orn, H. (1986b). Psychiatric disorders, spouse abuse and child abuse. *Acta Psychiatrica Belgica, 86*, 444–449.

Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology, 53*, 371–399.

Broidy, L. M., Nagin, D. S., Tremblay, R. E., Bates, J. E., Brame, B., Dodge, K. A., et al. (2003). Developmental trajectories of childhood disruptive behaviors and adolescent delinquency: A six-site, cross-national study. *Developmental Psychology, 39*, 222–245.

Brooks-Gunn, J., & Duncan, G. J. (1997). The effects of poverty on children. *Future of Children, 7*, 55–71.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods and Research, 21*, 230–258.

Burgess, R. L., & Conger, R. D. (1978). Family interaction in abusive, neglectful, and normal families. *Child Development, 49*, 1163–1173.

Chang, J. J., Halpern, C. T., & Kaufman, J. S. (2007). Maternal depressive symptoms, father's involvement, and the trajectories of problem behaviors in a US national sample. *Archives of Pediatrics and Adolescent Medicine, 161*, 697–703.

Cicchetti, D., & Rogosch, F. A. (1997). The role of self-organization in the promotion of resilience in maltreated children. *Development and Psychopathology, 9*, 797–815.

Costello, E. J., Compton, S. N., Keeler, G., & Angold, A. (2003). Relationships between poverty and psychopathology: A natural experiment. *Journal of the American Medical Association, 290*, 2023–2029.

Crockenburg, S. B., & Forgays, D. (1996). The role of emotion in children's understanding and emotional reactions to marital conflict. *Merrill-Palmer Quarterly, 42*, 22–47.

Cummings, E. M., Ballard, M., El-Sheikh, M., & Lake, M. (1991). Resolution and children's responses to inter-adult anger. *Developmental Psychology, 27*, 462–470.

Cummings, E. M., Vogel, D., Cummings, J. S., & El-Sheikh, M. (1989). Children's responses to different forms of conflict expression of anger between adults. *Child Development, 60*, 1392–1404.

Davies, P. T., & Cummings, E. M. (1998). Exploring children's emotional security as a mediator of the link between marital relations and child adjustment. *Child Development, 69*, 124–139.

Davies, P. T., Myers, R. L., Cummings, E. M., & Heindel, S. (1999). Adult conflict history and children's subsequent responses to conflict. *Journal of Family Psychology, 13*, 610–628.

DiLalla, L. F., & Gottesman, I. I. (1991). Biological and genetic contributions to violence: Widom's untold tale. *Psychological Bulletin, 109*, 125–129.

Dowd, K., Kinsey, S., Wheeless, S., Thissen, R., Richardson, J., Suresh, R., et al. (2004). *National Survey of Child and Adolescent Well-Being (NSCAW): Combined Waves 1-4 data user's manual*. Durham, NC: Research Triangle Institute.

Edleson, J. L. (1999). The overlap between child maltreatment and woman battering. *Violence Against Women, 5*, 134–154.

Egami, Y., Ford, D. E., Greenfield, S. F., & Crum, R. M. (1996). Psychiatric profile and sociodemographic characteristics of adults who report physically abusing or neglecting children. *American Journal of Psychiatry, 153*, 921–928.

Enders, C. K. (2001). The performance of the full information maximum likelihood estimator in multiple regression models with missing data. *Educational and Psychological Measurement, 61*, 713–740.

Farver, J. A. M., Xu, Y., Eppe, S., Fernandez, A., & Schwartz, D. (2005). Community violence, family conflict, and preschoolers' socioemotional functioning. *Developmental Psychology, 41*, 160–170.

Fox, N. A., & Leavitt, L. A. (1995). *The Violence Exposure Scale for Children (VEX)*. College Park: University of Maryland.

Hamagami, F., & McArdle, J. J. (2001). Advanced studies of individual differences linear dynamic models for longitudinal data analysis. In G. Marcoulides & R. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 203–246). Mahwah, NJ: Erlbaum.

Hay, T., & Jones, L. (1994). Societal interventions to prevent child abuse and neglect. *Child Welfare, 73*, 379–403.

Holden, G. W., & Ritchie, K. L. (1991). Linking extreme marital discord, child rearing, and child behavior problems: Evidence from battered women. *Child Development, 62*, 311–327.

Jaffee, S. R., Caspi, A., Moffitt, T. E., Taylor, A., & Arseneault, A. (2002). Influence of adult domestic violence on children's internalizing and externalizing problems: An environmentally informative twin study. *Journal of the American Academy of Child and Adolescent Psychiatry, 41*, 1095–1103.

Jaffee, S. R., Caspi, A., Moffitt, T. E., Taylor, A. & Arseneault, A. (2004). Physical maltreatment victim to antisocial child: Evidence of an environmentally mediated process. *Journal of Abnormal Psychology, 113*, 44–55.

Jaffee, S. R., & Gallop, R. (2007). Social, emotional, and academic competence among children who have had contact with child protective services: Prevalence and stability estimates. *Journal of the American Academy of Child & Adolescent Psychiatry, 46*, 757–765.

Jaffe, P. G., Wolfe, D. A., & Wilson, S. K. (1990). *Children of battered women*. Newbury Park, CA: Sage.

Jaffe, P., Wolfe, D., Wilson, S. K., & Zak, L. (1986). Family violence and child adjustment: A comparative analysis of girls' and boys' behavioral symptoms. *American Journal of Psychiatry, 143*, 74–76.

Jouriles, E. N., Murphy, C. M., & O'Leary, K. D. (1989). Interspousal aggression, marital discord, and child problems. *Journal of Consulting and Clinical Psychology, 57*, 453–455.

Kaplan, S. J., Pelcovitz, D., & Labruna, V. (1999). Child and adolescent abuse and neglect research: A review of the past 10 years. Part I: Physical and emotional abuse and neglect. *Journal of the American Academy of Child & Adolescent Psychiatry, 10*, 1214–1222.

Katz, L. F., & Gottman, J. M. (1997). Buffering children from marital conflict and dissolution. *Journal of Clinical Child Psychology, 26*, 157–171.

Kendall, P. C. (2000). *Childhood disorders*. East Sussex, UK: Psychology Press.

Kernic, M. A., Wolf, M. E., Holt, V. L., McKnight, B., Huebner, C. E., & Rivara, F. P. (2003). Behavioral problems among children whose mothers are abused by an intimate partner. *Child Abuse & Neglect, 27*, 1231–1246.

King, L. A., King, D. W., McArdle, J. J., Saxe, G. N., Doron-LaMarca, S., et al. (2006). Latent difference score approach to longitudinal trauma research. *Journal of Traumatic Stress, 19*, 771–785.

Little, R. J., & Rubin, D. B. (1987). *Statistical analyses with missing data*. New York: Wiley.

MacCallum, R. C. (1990). The need for alternative measures of fit in covariance structure modeling. *Multivariate Behavioral Research, 25*, 157–162.

Magdol, L., Moffitt, T. E., Caspi, A., & Silva, P. A. (1998). Developmental antecedents of partner abuse: A prospective-longitudinal study. *Journal of Abnormal Psychology, 107*, 375–389.

Margolin, G., & Gordis, E. B. (2000). The effects of family and community violence on children. *Annual Review of Psychology, 51*, 445–479.

McGuffin, P., Owen, M. J., & Gottesman, I. I. (Eds.). (2002). *Psychiatric genetics and genomics*. Oxford, UK: Oxford University Press.

McLoyd, V. C. (1997). The impact of poverty and low socioeconomic status on the socioemotional functioning of African-American children and adolescents: Mediating effects. In R. D Taylor & M. C. Wang (Eds.), *Social and emotional adjustment and family relations in ethnic minority families*. Mahwah, NJ: Lawrence Erlbaum.

Medsker, G. J., Williams, L. J., & Holahan, P. J. (1994). A review of current practices for evaluating causal models in organizational behavior and human resources management research. *Journal of Management, 20*, 439–464.

Moffitt, T. E., & Caspi, A. (1998). Annotation: Implications of violence between intimate partners for child psychologists and psychiatrists. *Journal of Child Psychology and Psychiatry, 39*, 137–144.

National Research Council (1993). *Understanding child abuse and neglect. Panel on Research on Child Abuse and Neglect. Commission on Behavioral and Social Sciences and Education*. Washington, DC: National Academy Press.

O'Brien, M., John, R. S., Margolin, G., & Erel, O. (1997). Reliability and diagnostic efficacy of parents' reports regarding children's exposure to marital aggression. *Violence and Victims, 9*, 45–52.

Pollak, S., Vardi, S., Bechner, A. M. P., & Curtin, J. (2005). Physically abused children's regulation on attention in response to hostility. *Child Development, 76*, 968–977.

Raviv, A., Erel, O., Fox, N. A., Leavitt, L. A., Raviv, A., Dar, I., et al. (2001). Individual measurement of exposure to everyday violence among elementary schoolchildren across various settings. *Journal of Community Psychology, 29*, 117–140.

Raviv, A., Raviv, A., Shimoni, H., Fox, N. A., & Leavitt, L. A. (1999). Children's self-report of exposure to violence and its relation to emotional distress. *Journal of Applied Developmental Psychology, 20*, 337–353.

Raykov, T. (2005). Analysis of longitudinal studies with missing data using covariance structure modeling with full-information maximum likelihood. *Structural Equation Modeling: A Multidisciplinary Journal, 12*, 493–505.

Shackman, J., Shackman, A., & Pollak, S. D. (2007). Physical abuse amplifies attention to threat and increases anxiety in children. *Emotion, 7*, 838–852.

Shahinfar, A., Fox, N. A., & Leavitt, L. A. (2000). Preschool children's exposure to violence: Relation of behavior problems to parent and child reports. *American Journal of Orthopsychiatry, 70*, 115–125.

Slep, A. M. S., & O'Leary, S. G. (2001). Examining partner and child abuse: Are we ready for a more integrated approach to family violence? *Clinical Child and Family Psychology Review, 4*, 87–107.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 25*, 173–180.

Straus, M. A. (1992). Children as witness to marital violence: A risk factor for lifelong problems among a nationally representative sample of American men and women. In D. F. Schwarz (Ed.), *Abused and battered: Social and legal responses to family violence*. New York: Aldine de Gruyter.

Straus, M. A., Hamby, S. L., Finkelhor, D., Moore, D. W., & Runyan, D. (1998). Identification of child maltreatment with the Parent-Conflict Tactics Scales (CTS2): Development and preliminary psychometric data. *Journal of Family Issues, 17*, 283–316.

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin, 133*, 859–883.

Trickett, P. K., & Kuczynski, L. (1986). Children's misbehaviors and parental discipline strategies in abusive and nonabusive families. *Developmental Psychology, 22*, 115–123.

Trickett, P. K., & McBride-Chang, C. (1995). The developmental impact of different forms of child abuse and neglect. *Developmental Review, 15*, 311–337.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38*, 1–10.

U.S. Department of Health and Human Services, Administration for Children and Families, Children's Bureau. (2005). *Child maltreatment 2003*. Washington, DC: U.S. Government Printing Office, Author. Retrieved January 25, 2006, from http://www.acf.hhs.gov/programs/cb/pubs/cm03/index.htm

Vasta, R. (1982). Physical child abuse: A dual-component analysis. *Developmental Review, 2*, 125–149.

Ware, J., Kosinski, M., & Keller, S. (1998). *SF-12: How to score the SF-12 Physical and Mental Health Summary Scales* (3rd ed.). Lincoln, RI: Quality Metric.

Ware, J. E., Jr., Kosinski, M., & Keller, S. D. (1996). A 12-item short-form health survey: Construction of scales and preliminary tests of reliability and validity. *Medical Care, 34*, 220–233.

Wolfe, D. A., Crooks, C. V., Lee, V., McIntyre-Smith, A., & Jaffe, P. G. (2003). The effects of children's exposure to domestic violence: A meta-analysis and critique. *Clinical Child and Family Psychology Review, 6*, 171–187.

Wolfe, D. A., Jaffe, P., Wilson, S. K., & Zak, L. (1985). Children of battered women: The relation of child behavior to family violence and maternal stress. *Journal of Consulting and Clinical Psychology, 53*, 657–665.

Child Development, September/October 2008, Volume 79, Number 5, Pages 1513–1523

Limitations in 4-Year-Old Children's Sensitivity to the Spacing Among Facial Features

Catherine J. Mondloch and Kendra Thomson
Brock University

Four-year-olds' sensitivity to differences among faces in the spacing of features was tested under 4 task conditions: judging distinctiveness when the external contour was visible and when it was occluded, simultaneous match-to-sample, and recognizing the face of a friend. In each task, the foil differed only in the spacing of features, and spacing alterations were within normal limits. Children performed at chance levels in all but 1 task—match-to-sample, and in that task, only 10 of 18 children were correct on more than 50% of the trials. Sensitivity to the spacing of facial features in identity and distinctiveness tasks is very weak at 4 years of age—at least when the changes do not exceed ±2.5 *SD* of normal variability.

Adults are experts at recognizing facial identity: They can recognize thousands of faces, including, at least for a short time, ones they encounter briefly in the course of a laboratory experiment (see Bruce & Young, 1998, for a review). Adults use multiple cues to recognize the identity of individual faces, two of which have been studied extensively in children: the shape of individual features (e.g., eyes, mouth, and chin: *featural processing*) and the spacing of the facial features (e.g., the distance between the eyes, a type of configural processing called *second-order relational processing*; for a review, see Maurer, Le Grand, & Mondloch, 2002). Despite the early emergence of some face-processing skills, adult-like expertise is not achieved until some time during adolescence: Even in matching tasks, which eliminate memory demands, performance improves dramatically between 4 and 11 years of age (Bruce et al., 2000; see also De Sonneville et al., 2002).

Two lines of evidence suggest that children's poor performance on face recognition tasks can be attributed in part to immature sensitivity to second-order relations. First, children perform like adults on identity tasks that tap sensitivity to the appearance of individual features. For example, when asked to make same/different judgments about pairs of faces that differ only in the shape of the external contour, 6-year-olds are as accurate as adults. They are nearly like adults when making same/different judgments about pairs of faces that differ only in the shape of internal features (eyes and mouth), with no statistical difference by 10 years of age (Mondloch, Le Grand, & Maurer, 2002; see also Freire & Lee, 2001). Second, children aged 6–14 years make significantly more errors than adults when asked to make same/different judgments about pairs of faces that differ only in the spacing of features (Freire & Lee, 2001; Mondloch, Dobson, Parsons, & Maurer, 2004; Mondloch et al., 2002, 2003).

There is general consensus that children are less sensitive than adults to the spacing among facial features (but see Gilchrist & McKone, 2003; Pellicano, Rhodes, & Peters, 2006), but there is debate as to when this sensitivity first emerges. Although less accurate than adults, even 6-year-olds' accuracy is above chance levels in making same/different judgments about upright faces differing only in the spacing of their features ($M = 0.69$ correct vs. 0.82 correct for adults tested with the same task; Mondloch et al., 2002). In addition, like adults, 7-year-old children remember distinctive faces better than typical faces, regardless of whether the faces are made distinctive by manipulating features (e.g., bushy eyebrows) or their spacing (e.g., mouth up; Gilchrist & McKone, 2003).

To investigate when sensitivity to the spacing among facial features first emerges, Mondloch, Leis, and Maurer (2006) asked 4-year-old children to recognize each of three faces when paired with a foil that

---

This research was funded by an NSERC Discovery Grant to Catherine J. Mondloch. The original adult faces were taken from the MacBrain Face Stimulus Set. Development of the MacBrain Face Stimulus Set was overseen by Nim Tottenham and supported by the John D. and Catherine T. MacArthur Foundation Research Network on Early Experience and Brain Development. Please contact Nim Tottenham at tott0006@tc.umn.edu for more information concerning the stimulus set.

Correspondence concerning this article should be addressed to Catherine J. Mondloch, Department of Psychology, Brock University, St. Catharines, Ontario, Canada L2S 3A1. Electronic mail may be sent to cmondloch@brocku.ca.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7905-0019

differed from the target in one of three ways: the shape of the external contour, the appearance of individual features, or the spacing among features. Four-year-olds were familiarized with two of the target faces over a 2-week period by providing them with a storybook depicting two children on a visit to the farm. To encourage attention to second-order relations, Mondloch et al. varied the appearance of facial features by presenting the faces of the two children in the storybook from different points of view and with a variety of emotional expressions. The third target face was the child's own. Recognition was tested in the context of a computerized game in which 4-year-olds used a "magic wand" to put children on a bus or train during a trip to the zoo. Four-year-olds made no errors on control trials in which the foil was the face of another child or in which the foil had a different external contour; they performed above chance levels when the foil differed in the shape and color of the eyes and mouth. In contrast, 4-year-olds performed at chance levels when the foil differed from the target only in the spacing of facial features, both when tested with the two target faces from the storybook and when tested with their own face. Mondloch et al. concluded that sensitivity to second-order relations emerges after 4 years of age—at least for facial identity tasks when the spacing of features remains within normal limits

In the same year that Mondloch et al. (2006) published their manuscript, two other studies reported that 4-year-old children are sensitive to the spacing of facial features. Pellicano et al. (2006) tested children's sensitivity to the spacing of facial features using the *whole/part* paradigm. The *whole/part effect* is a measure of holistic processing (Tanaka & Farah, 1993) and refers to the finding that adults and children as young as 4 years of age recognize the features from an individual's face more easily in the context of the whole face (e.g., Larry's nose in Larry's face) than in isolation (Pellicano & Rhodes, 2003; Tanaka, Kay, Grinnell, Stansfield, & Szechter, 1998). The usual interpretation of the whole/part paradigm is that poor performance in the isolated features condition indicates that faces are processed holistically (i.e., as a gestalt) and that characteristics of individual features are encoded in the context of this gestalt representation (Tanaka & Farah, 2003). Disrupting holistic processing by presenting features in isolation impairs recognition of features, as does altering the gestalt by changing second-order relations (Tanaka & Sengo, 1997). Pellicano et al. showed that altering the spacing of facial features in one part of a face also impairs the ability of 4-year-olds to recognize facial features in the whole condition.

McKone and Boyer (2006) reported that 4-year-olds are sensitive to differences in the spacing of facial features when asked to judge which of two faces is more distinct. They manipulated facial distinctiveness by altering the appearance of individual features (e.g., by making the eyebrows bushy) or by altering the spacing among features. Based on adult ratings of the original and distorted faces viewed one at a time, McKone and Boyer created pairs of faces in which differences in distinctiveness between the original face and its altered version were large, medium, or small. Adults and 4-year-olds were then asked to indicate which member of each pair was more distinctive; adults' performance was above chance for all pairings, and 4-year-olds' performance was above chance when differences in distinctiveness were either large or medium.

The purpose of the present set of studies was to resolve these discrepant findings. It may be that sensitivity to differences among faces in the spacing of facial features emerges at different rates for recognizing a face's identity than for judging distinctiveness. In the current series of studies, we measured 4-year-olds' sensitivity to the spacing of facial features under several task conditions—when judging distinctiveness, in a simultaneous match-to-sample task, and when recognizing the faces of friends from a day care. In all these studies, the spacing changes were identical to those used by Mondloch et al. (2006). This allowed us to assess whether 4-year-olds would show sensitivity to those spacing changes under different task conditions. Furthermore, we made each task child friendly: The number of trials was minimized (four test trials plus eight control trials), and children were presented with a "game" in which they used a magic wand to put faces on a bus or put twins into a "twin clubhouse."

Experiment 1a

Previously, we reported that 4-year-olds were not able to recognize either faces from a storybook or their own face when targets were paired with foils that differed only in the spacing of facial features (Mondloch et al., 2006). This result is in contrast to that of McKone and Boyer (2006) who reported that 4-year-olds were able to select the more distinctive of two faces that differed in the spacing of features, as long as the differences in distinctiveness were either large or medium according to adult judgments. In Experiment 1a, we showed children four pairs of faces that differed in the spacing of features and asked them to use a magic wand to put the more unusual child on an "unusual" bus. Each of the spatially altered faces had been rated as more distinctive than its original version when adults viewed three versions of each of 26 faces sequentially (Mondloch et al., 2006); in that study, 4-year-olds performed at chance levels in an identity task. If the spacing of facial features influences children's judgments of distinctiveness by 4 years of age, they should select the more distinctive version of each face (Mondloch et al., 2006). We also tested a group of adults in the current study using the method designed for children. We expected that adults would perform without error, thus validating our stimulus set.

Method

Participants

Eighteen 4-year-old children (9 females; mean age = 4.43 years, range = 4.0–4.9 years) participated in this study. Twelve adults (mean age = 19.9 years, range = 19.0–23.6) with normal or corrected-to-normal vision served as a control group. All participants were Caucasian. Children in all the experiments were recruited from a database of children whose parents had volunteered for participation or from local day cares whose parents had provided informed consent. Adult participants received course credit.

Materials

Test stimuli. On each of the four test trials, we presented an original (i.e., unaltered) face paired with a foil that differed only in the spacing of internal features. We selected four face pairs from Experiment 1 of Mondloch et al. (2006). We used the original faces of the two storybook characters plus the faces from two of the participants in that study. Adults had previously rated each of these faces on a distinctiveness scale (Mondloch et al., 2006, Experiment 2b) and reported that the spatially altered versions were more distinct than the original faces (mean difference = 1.66, *SE* = 0.22). In this and all subsequent studies, the faces were close to life size: The distance between the chin and the hairline was 130 mm. Based on anthropomorphic norms (Farkas, 1994), we had moved the eyes up/down 21 pixels (where 1 pixel = 0.353 mm; 1.3 *SD*; 0.71° from the testing distance of 60 cm) or closer together or farther apart 14 pixels (2.8 *SD*; 0.47° from 60 cm) and we moved the mouth up/down 7 pixels (0.8 *SD*; 0.24° from 60 cm). Thus, these spacing changes covered most of the natural variability in the real world and consequently remained within normal limits. Two of the test faces had the eyes and mouth moved up and two had the eyes and mouth moved down. Figure 1a provides an example.

Criterion stimuli. Two sets of criterion stimuli were used to ensure that the 4-year-olds understood the task and that they continued to "play the game" throughout the procedure. One set consisted of four animals (e.g., a horse, a squirrel) paired with a "weird" version of that same animal (e.g., a horse with stripes, a squirrel with pigs' ears; Figure 1b). The other set consisted of grayscale photographs of adult Caucasian women paired with a spatially altered version in which the spacing of facial features was outside normal limits. Adults and 8-year-olds rate these spatially altered versions as grotesque (Mondloch et al., 2004), and unpublished data of Mondloch and Maurer (2003) show that 4-year-olds find these faces weird (Figure 1c).

Procedure

The child sat 60 cm away from a 58-cm computer screen. We began by presenting a picture of two buses, one of which was a typical school bus and one of which was weird. The experimenter said, "Look at these two buses. Do you think that one of these buses is a bit strange or unusual or weird? Which one? Why do you think that bus is weird? Good job! Now we are going to go for a ride with this bus and help some people and animals get back to their town. Do you think you can help me? Great!" Each trial consisted of two photos of animals, children's faces, or adult female faces. The experimenter asked, "Do you think one of these (animals, females, or children) is a bit strange, unusual, weird, or might stick out to you if you saw them on the playground? Good, now wave your magic wand to put the weird (animal, female, or child) on the weird bus!" After the child waved their magic wand over one of the pictures, a small picture of that stimulus appeared on the bus. The order of the trials was such that all participants received four criterion trials (two animals, two adult female faces), followed by two test trials (children's faces). Participants then received two more criterion trials (animals), followed by the remaining two test trials. We ended with two final criterion trials (adult female faces) to ensure that children were still "playing the game." An alternate version of the task was created so that the left–right positions of the stimuli were reversed. To ensure that the experimenter was blind to the location of the more distinctive stimulus, another laboratory member opened one of the two files on the computer for the experimenter. The experimenter viewed the stimuli on a laptop computer on which the pictures were occluded. This

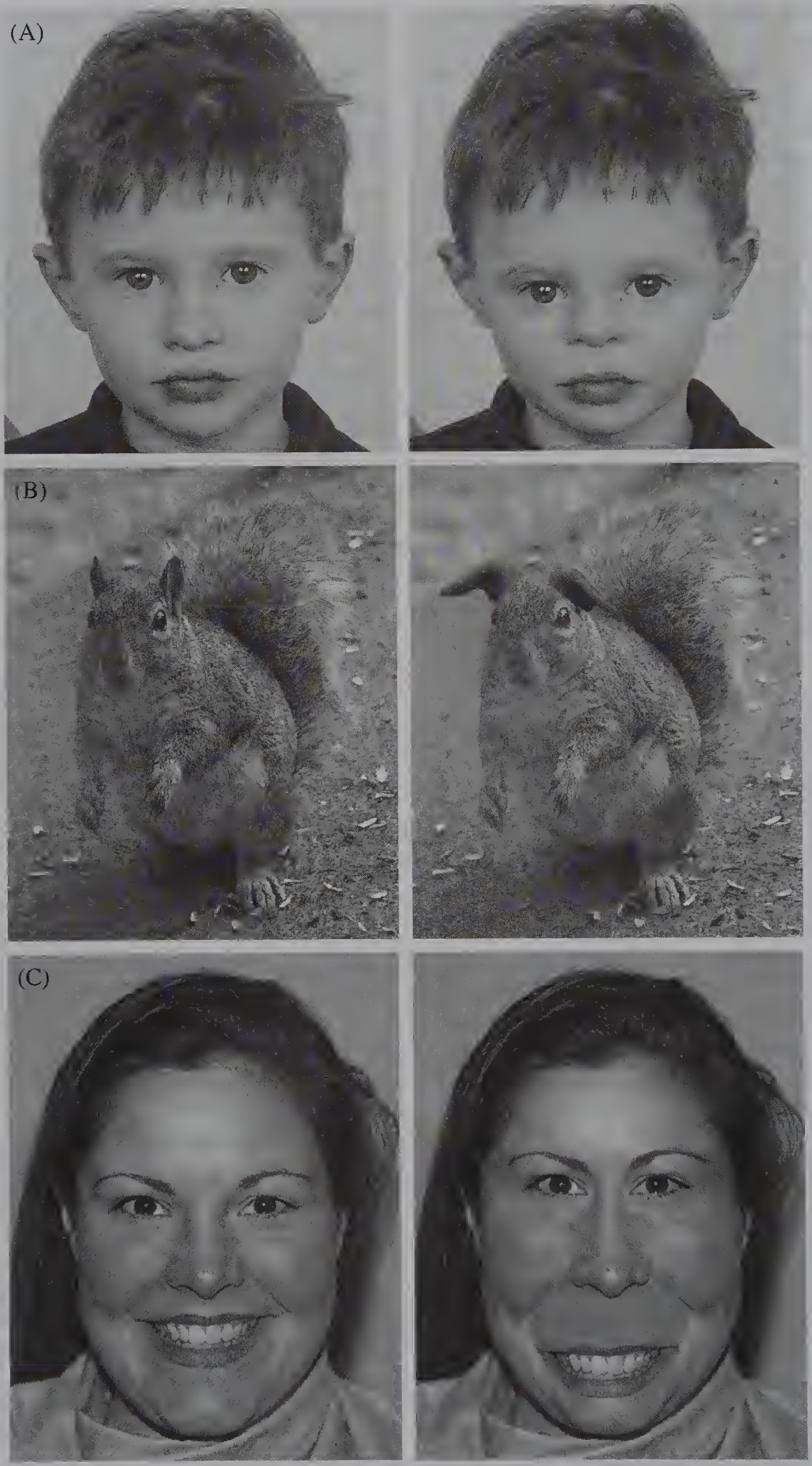


Figure 1. Examples of stimuli used during (a) test trials, (b) criterion trials in which animals were presented, and (c) criterion trials in which adult faces were presented. In each case, the original picture is shown on the left and the altered (i.e., distinctive) picture is shown on the right. The spacing changes made to adult faces were outside normal limits; spacing changes made to children's faces were within normal limits.

protocol was approved by the Research Ethics Board at Brock University.

Results and Discussion

Children made almost no errors when asked to find the more distinctive animal, mean correct (on four trials) = 3.90, $t(17) = 35.0$, $p < .001$, and when asked to find the more distinctive adult face, mean correct = 3.70, $t(17) = 14.6$, $p < .001$ (all tests were one-tailed). Children performed at chance levels, mean correct = 1.90, on trials in which the original version of a child's face was paired with its spatially altered version and the spacing changes were within normal limits. Only 7 of the 18 children tested were correct on at least three of the four test trials. A control group of 12 adults (7 females) performed at 100% accuracy on all trial types, confirming that the spatially altered faces are easily detected as the most distinctive by adults. These data suggest that 4-year-olds are not able to use information about the spacing of facial features to determine which of two faces is more distinctive when the spacing changes are within normal limits. This is in contrast to the findings of McKone and Boyer (2006) who reported above-chance performance in 4-year-olds—at least when differences in distinctiveness were large or medium in size. We note, however, that McKone and Boyer occluded the external contour of their face stimuli, whereas the hair was visible in our stimuli, albeit identical within each stimulus pair. Several studies have demonstrated that the external contour is a salient stimulus to young children (Campbell & Tuck, 1995; Campbell, Walker, & Baron-Cohen, 1995; Mondloch et al., 2006); to see whether the presence of hair interfered with children's performance, we replicated the procedure of Experiment 1a, but occluded the hair.

Experiment 1b

Method

Participants

Eighteen 4-year-old children (5 females; mean age = 4.38 years, range = 4.0–4.7 years) participated in this study. All participants were Caucasian. Experiment 1b was identical to Experiment 1a, with one exception: The hair was occluded on children's faces (participants were told that the children were peeking through windows).

Results and Discussion

No child made an error on criterion trials, but performance was at chance levels when children were asked to find the more distinctive face on test trials, mean correct = 2.20, $t(17) = 0.62$, $p > .2$. Only 6 of the 18 participants were correct on at least three of the four test trials. These data confirm that 4-year-olds do not attend to the spacing among facial features per se when making distinctiveness judgments.

At first glance, it appears that our results contradict those of McKone and Boyer (2006) who reported that 4-year-old children judge spatially altered faces as more distinctive than their original versions. However, in their study, 4-year-olds were much less accurate than adults even when the spacing changes were large (mean correct = 0.90 vs. 0.62 for adults and children, respectively). When differences in distinctiveness were small, 4-year-olds did not perform above chance levels, whereas adults were correct on approximately 75% of the trials. Thus, even in McKone and Boyer's study, 4-year-olds' sensitivity to the spacing of facial features was much weaker than that of adults and was most reliable when differences in distinctiveness were large according to adult ratings. Likewise, in the current study, 4-year-olds were able to use information about the spacing of facial features when the spacing changes were large and outside normal limits (i.e., when making distinctiveness judgments about the adult faces). The high level of accuracy shown for adult faces with large spatial distortions (12 children correct on four of four trials; 6 children correct on three of four trials) in contrast to chance performance for children's faces with spatial distortions that contrast more or less average spacing with the edge of natural limits suggests that 4-year-old children are sensitive to differences among faces in the spacing of features but only when those differences are large. One potential limitation of our study is that two training trials in which differences in distinctiveness were large preceded the first two test trials. We did this in order to familiarize children with the task. Although one possibility is that the spacing changes made to the children's faces were harder to detect after viewing the large spacing changes in adult faces, we reasoned that viewing the large spacing may direct 4-year-olds' attention to this cue to distinctiveness. Nonetheless, we note that this may have made the task more difficult for the children.

Experiment 2

In Experiment 2, we simplified the task for 4-year-old children by presenting them with a simultaneous match-to-sample task. On each of four test trials, they were shown a child's face (either original or spatially altered) and then shown a pair of faces, one of which was identical to the target and one of which differed only in the spacing of features. Children were asked which member of the face pair matched the target; they then placed the "twins" in a "twin clubhouse." This task involved no memory demands and allowed the child to compare the target face to two test faces with no time restrictions. We also included criterion trials in which children were presented with sets of adult faces and animals from Experiment 1.

Method

Participants and Materials

Eighteen 4-year-old children (8 females; mean age = 4.44 years, range = 4.0–4.9 years) participated in this study. All participants were Caucasian. The test stimuli were identical to those used in Experiment 1a; they were printed and laminated to allow the child to handle the faces. We created a twin clubhouse into which children were invited to place identical pictures.

Procedure

Each participant sat at a table with the stack of cards containing stimulus pairs in front of him or her. The experimenter sat next to the participant with an additional stack of individual stimuli. The twin clubhouse was placed directly in front of the participants. First, participants were shown a drawing of three bears, two of which were identical and one that was only slightly different. The experimenter said, "Do you know what twins are? Can you pick out the twins from these three bears?" If participants did not understand the differences between the bears, the experimenter explained, "Twins are people or animals that look exactly the same." Next, the experimenter told participants, "Today I have a special job for you! You are going to be a 'twin detective'! Some kids, animals and adults are starting a special 'twin clubhouse' where only twins can go. Can you help me find twins to put in the twin clubhouse? Great! First I am going to flip over a card with an animal (or face) on it and then you can flip your card that will have two pictures of animals (or faces) on it. Your job is to pick the animal (or face) from your card that is exactly like the animal (or face) on my card (the animal's/person's twin)! Then we can put the twin animals/people in the twin clubhouse. Remember to look carefully at both of the pictures on your card!"

Results and Discussion

Children made no errors on the animal criterion trials, and only 3 children made a single error on the criterion trials in which we presented adult faces. Their performance was above chance, $M = 2.40$, $t(17) = 1.80$, $p < .05$, when tested with children's faces, though only 10 of the 18 participants were correct on at least three of the four test trials. These data suggest that the visual system of 4-year-old children is able to detect differences in the spacing of facial features, a finding that is consistent with report by Pellicano et al. (2006) of sensitivity to the spacing of features in the whole/part paradigm. Four-year-olds showed sensitivity to the spacing of facial features in our match-to-sample task in which they were required only to match facial identity; this is in contrast to when they were asked to recognize a face from memory (Mondloch et al., 2006) or to decide which of two faces was distinctive (Experiment 1). Nonetheless 8 of the 18 children tested were not correct on more than 50% of the trials, suggesting that 4-year-olds' sensitivity to the spacing of facial features is weak even in a matching task. Based on this, it is unlikely that they use the spacing of facial features to recognize faces in the real world.

Nonetheless, we may have underestimated children's sensitivity to second-order relations by testing them with unfamiliar faces. Carver et al. (2003) reported an effect of familiarity on the Nc component of the event-related potential waveform between 18 and 24 months of age; this raises the possibility that children may process highly familiar faces differently from unfamiliar faces. Mondloch et al. (2006) tested 4-year-olds with familiarized faces from a storybook and with their own faces. We reasoned, however, that 4-year-olds do not see their own face as often as that of people with whom they interact on a daily basis and they do not see their own face from a variety of perspectives and displaying a range of facial expressions. In Experiment 3, we tested the hypothesis that 4-year-olds may be sensitive to the spacing of facial features when asked to identify the "real" picture of a highly familiar face—that of a friend from their day care. To do so, we replicated the original method of Mondloch et al. (2006) but replaced pictures of the storybook character and the participant's own face with the faces of three friends from the participant's day care.

Experiment 3

Method

Participants

Eighteen 4-year-old children (6 females; mean age = 4.25 years, range = 4.0–4.9 years) participated in this study. All participants were Caucasian. All participants attended a day care and were highly familiar with the faces of children who served as models. Two additional children were tested but excluded from the analyses. One child was able to identify only two of his three friends when shown the initial array of stimuli, and the other was incorrect on all three contour trials. This poor performance is very atypical (Mondloch et al., 2006); therefore, we excluded these children to avoid underestimating 4-year-olds' sensitivity to the spacing of features.

Stimuli

For each participant, a day-care worker identified three children with whom the participant was highly familiar; these children served as models. After obtaining informed parental consent, we took three pictures of each model's face. We then created new versions of each of these pictures. On contour trials, we presented one original version paired with a foil in which we combined the external contour (hair, ears, and chin) of a different child with the internal features and the clothing of the original image. On test trials, we presented an original photograph paired with a spatially altered version. These spatial alterations were identical to those used by Mondloch et al. (2006). On half of the trials, the eyes and mouth were moved either up or down (as in Experiments 1 and 2); on the remaining trials, the eyes were moved 14 pixels closer together and the mouth was moved down 7 pixels or the eyes were moved 14 pixels farther apart and the mouth was moved up 7 pixels. A different original picture was presented on each spacing trial to ensure that participants were recognizing a face, rather than a particular photograph. On control trials, we presented one of the original pictures of the model paired with a picture of another 4-year-old child with whom the participant was not familiar.

Procedure

The child sat 60 cm away from a 58-cm computer screen; during the test, the experimenter faced the child so as to be blind as to on which side of the monitor the "correct" face was presented. We began by presenting an array of six pictures, each of which depicted a 4-year-old child and three of which were friends of the participant. The participant was asked to identify their three friends.

On the first three test trials, the child was shown one of the original faces paired with a face with the same internal features but a different external contour. The background picture consisted of a school bus. The child was told, "Today we are going to pretend that some of your friends are going on a trip to the zoo! But the bus driver doesn't know what all the kids in the class look like, so maybe you can help him out. Here is a magic wand and when you wave the wand the children will get on the bus." Two faces appeared on the face at this time. "Look! It's 'Sam'! And there's another boy who looks a little bit like Sam. Can you wave the magic wand over the real picture of Sam to put him on the bus?" After the child pointed the magic wand to the face he or she thought was Sam's, a new display appeared that consisted of the interior of the bus and a small picture of Sam's face. On this and every subsequent trial, the correct response was placed on the bus (or train). Two additional contour trials followed in which the child was asked to help the bus driver by pointing to their two remaining friends. The purpose of these trials was to illustrate to the child that the members of each face pair would be quite similar, but that only one would be the real person.

The child was then shown a series of four animal pictures (e.g., a monkey, an elephant) and the experimenter continued the story about the trip to the zoo. Then, the game continued: "Oh look, there is a train ride. The kids want to go on the train ride at the zoo, but the conductor doesn't know what the kids look like. Maybe you can help him out. But, it is going to be harder now. Just for today, Sam has a brother who looks a lot like him so it's going to be more difficult to tell them apart so look very carefully." Two faces were then presented on the screen; the faces differed only in the spacing of the facial features. "Here's Sam and Sam's brother. They look a lot alike, but only one of them is really Sam. Can you wave your magic wand at Sam to put him on the train?" When the child waved the magic wand, the experimenter presented a new picture that consisted of a picture of the train overlaid with a small picture of Sam's face. Two additional trials followed in which different friend's faces were presented. The child was then told that it was time for the children to go home but that the bus driver was sad because he could not remember what the children in the class looked like. A second set of spacing trials then ensued, one for each of the three original faces.

To ensure that the child was still attentive at the end of the procedure, three control trials were included. The child was told, "So now the bus takes them all the way back to school. At school, their mommies and daddies are waiting to pick them up. Can you help Sam's mom pick out Sam?" Two faces were presented side-by-side on the screen, but this time, Sam was presented next to a completely different boy. The background consisted of a cartoon version of a car.

After the child pointed to the face he or she thought belonged to Sam, Sam appeared in the passenger side of the car. The same test was then repeated for the other two friends. Any child who could not correctly identify these faces would have been excluded from the study. However, not one child made an error.

For each participant, we created two versions of the test in which the positions of the faces were reversed left to right. We then asked a laboratory member other than the experimenter to randomly open one of the two files to ensure that the experimenter was unaware of on which side the target face was presented.

Results and Discussion

One child made a single error on the contour trials, and no child made an error on control trials. In contrast, 4-year-olds performed at chance levels on the six spacing trials, $M = 3.39$, $t(17) = 1.13$, $p > .10$. Nine of the 18 participants were correct on at least four of the six test trials. Accuracy on the three trials in which the eyes were moved up and down ($M = 1.67$) did not differ from accuracy in which the eyes were moved horizontally, $M = 1.72$, $t(17) = 0.164$, $p > .50$. These results suggest that in previous studies, we had not underestimated 4-year-olds' sensitivity to the spacing among facial features by testing them with unfamiliar or familiarized faces. Even when tested with the faces of children with whom they interact on a daily basis, 4-year-old children do not use differences in the spacing of facial features for face recognition.

General Discussion

This set of studies indicates that 4-year-old children are not sensitive to small differences among faces in the spacing of facial features. They did not select spatially altered faces as more distinctive than the unaltered versions (Experiments 1a and 1b), and they did not select the unaltered versions as the real picture of their friends (Experiment 3). Across these three tasks, no more than half of the children performed better than a chance level of 50%. Their performance was above chance in the simultaneous match-to-sample task (Experiment 2), but only 10 of the 18 children were correct on more than 50% of the trials. Poor performance on these tasks cannot be attributed to 4-year-olds not understanding task instructions; their performance was almost errorless on control trials in which they were asked to select the more distinctive animal or adult female face in which the spacing changes were large and outside normal limits (Experiments 1a and 1b) and when asked to match the identity of these stimuli (Experiment 2). Similarly, only one child made a single error when asked to find the real picture of a friend's face when the foil differed in the appearance of the external contour. Poor performance is also unlikely due to the limited number of test trials. Our decision to include four test trials (plus eight control trials) in Experiments 1 and 2 was based on extensive pilot work showing that increasing the number of trials on a difficult task results in several children failing to complete the procedure and in random performance on control trials. A trial-by-trial analysis reveals no consistent improvement across trials. Across the first three studies, 27 of 54 children were correct on the first test trial and 31 of 54 children were correct on the last test trial. Rather, these results suggest that 4-year-old children are not sensitive to differences among faces in the spacing of facial features, at least when the spacing of features does not exceed ±2.5 *SD* of normal variability.

The nature of our spacing changes should have facilitated 4-year-olds' ability to use second-order relations in each task. In our distinctiveness tasks and in the match-to-sample task, we always moved the eyes vertically. This decision was based on adult ratings of distinctiveness from a previous study (Mondloch et al., 2006). Goffaux and Rossion (2007) reported that inversion disrupts the ability to discriminate faces in which features have been moved vertically more than the ability to discriminate faces in which features have been moved horizontally or in which the appearance of individual features has been altered. They suggest, "It may be that vertical relations are more significant for upright face discrimination than horizontal relations" (p. 1001). This is consistent with evidence from the composite face effect of stronger perceptual integration across the top and bottom halves of faces than across the two sides (Hole, 1994). Thus, the use of vertical movements should have made this task easier for children. In the friends study (Experiment 3), we moved the eyes vertically in half of the trials and horizontally in the other half. When we moved the eyes farther apart, the mouth was moved up; when we moved the eyes closer together, the mouth was moved down. These combinations of movement should have maximally disrupted the aspect ratio of the triangular relations in the original faces (Barton, Zhao, & Keenan, 2003), thus increasing 4-year-olds' sensitivity to the changes in spacing.

Our results are different from those of two previous studies reporting that 4-year-old children are sensitive to small differences among faces in the spacing of facial features. McKone and Boyer (2006) reported that 4-year-olds were able to select the more distinctive of two faces that differed only in the spacing of features, at least when differences in distinctiveness were medium or large. To create three face sets, they altered the spacing among features and then used adult ratings of distinctiveness to form face pairs in which *differences* in distinctiveness were classified as small, medium, or large. After completing our studies, McKone provided us with the complete set of the stimuli used by McKone and Boyer. We resized McKone and Boyer's stimuli to match ours and then measured how far features were moved when creating the distinctive face. We note several differences between our stimuli and those of McKone and Boyer. First, McKone and Boyer moved the eyes of all but two faces horizontally; the eyes of some faces also were moved vertically. Second, the mean number of pixels by which eyes were moved was larger in all three face sets (small: $M = 16.73$, medium: 15.81, large: 19.38) relative to the horizontal movements used in our current studies (14 pixels). The mean number of pixels by which the mouth was moved was also larger in the McKone and Boyer face sets (small: $M = 8.91$, medium: 8.76, large: 8.13) than in the current study (7 pixels). Third, within each face set, the number of pixels by which features were moved varied (e.g., eyes in the "small differences" face set were moved between 0 and 28 pixels), and within each set, the eyes of some faces were moved at least 20 pixels (4 *SD*). In sum, the average horizontal eye movements used by McKone and Boyer were larger than ours, and some were especially large. We note that despite using larger spacing changes, 4-year-olds' accuracy was low in the McKone and Boyer's study for the large (mean correct = 62%) and medium ($M = 58\%$) face sets and at chance (50%) for the small face set.

The results of our present studies also appear to be inconsistent with those of Pellicano et al. (2006) who showed that altering the spacing of facial features in one part of a face (e.g., the eyes) impairs the ability of 4-year-olds to recognize facial features in the same region or in another part of the face (e.g., the mouth). One possibility is that different task demands lead to differences in children's sensitivity to the spacing among features. In addition, the spacing among facial features in the part/whole task were larger than those used in the current study. Pellicano et al. created two versions of each face (e.g., one version with the eyes moved out and one with the eyes moved in). Those spacing changes were within normal limits and were identical to those used in Experiment 3 of the current study when features were moved horizontally. However, Pellicano et al. did not present the original face on test trials; rather, the model face had eyes moved in one direction (e.g., farther apart) and the test face had eyes moved in the opposite direction (i.e., closer together; L. Pellicano & G. Rhodes, personal communication, November 2007). This is in contrast to our studies in which the original, unaltered version of a face was presented with a spatially altered version in which the eyes were moved in one direction. Consequently, differences in the spacing among facial features were twice as large in the study by Pellicano et al. (approximately 5 *SD*) than in the current study (L. Pellicano & G. Rhodes, personal communication, November 2007).

Our current results also may be surprising in light of recent evidence that 5-month-old infants show sensitivity to differences in the spacing of facial features that fall within normal limits (Hayden, Bhatt, Reed, Corbly, & Joseph, 2007). Hayden et al. (2007) used a visual expectancy procedure in which two stimuli are presented on a screen. The stimuli are stationary for a period of time and then one of the stimuli begins to move; during the first part of the trial, infants look longer at the stimulus that they predict will not move, presumably because they look longer at the moving stimulus during the last part of the trial. Evidence for 7- and 5-month-old infants being sensitive to differences among faces in the spacing of features comes from their successfully predicting which of two faces would move when the pair differed only in the spacing among facial features. Hayden et al. concluded that sensitivity to the spacing among facial features emerges during infancy and they suggested that the 4-year-olds in our previous study in which children were tested with their own face and with faces familiarized via a storybook may have failed because of memory load. This explanation seems unlikely in the current series of studies, however, because there was no memory load either in our distinctiveness tasks or in our matching task. It is possible that 4-year-olds' poor performance when asked to recognize their own face, faces from a storybook (Mondloch et al., 2006), or the face of a close friend (Experiment 3), is due, in part, to memory load. This is of great interest, however, given adults' exquisite memory for the spacing of facial features in highly familiar faces (Ge, Luo, Nishimura, & Lee, 2003); when shown various versions of Chairman Mao's face that differed only in the spacing of features, Mainland Chinese adults were extremely sensitive to small alterations. Their accuracy in reporting which photographs were authentic matched that of Eastern Asian adults who were asked to make same/different judgments about pairs of stimuli from the same set. Nonetheless, memory load cannot completely explain children's poor performance in this series of tasks.

We reconcile our results with those of Hayden et al. (2007) in the following way. The visual system of infants is able to detect differences in the spacing of facial features and infants are able to use that information to predict which of two stimuli is about to move. Presumably, then, the visual system of 4-year-old children is able to detect differences in the spacing of facial features, but 4-year-olds do not use that information to judge distinctiveness or identity even when the spacing of facial features deviates from a more or less average position by up to 3 *SD*—a deviation that samples most of the variability among the faces 4-year-olds will attempt to identify in the real world. We agree with McKone and Boyer (2006) that 4-year-olds' judgments of distinctiveness can be influenced by the spacing among features but only when the spacing changes are quite large. This conclusion is consistent with 4-year-olds' performing well on our criterion trials in which we presented adult female faces and with previous work showing that 8-year-olds' rate these adult faces as grotesque, but their ratings are much lower than those of adults (Mondloch et al., 2004).

Overall, our results provide additional evidence that adult-like sensitivity to small differences among faces in the spacing of features is slow to develop. The ultimate development of sensitivity to the spacing of features is dependent upon visual input during early infancy (Le Grand, Mondloch, Maurer, & Brent, 2001, 2003), but sensitivity to spacing in identity and distinctiveness tasks emerges after 4 years of age—at least when spacing changes are ≤ 2.5 *SD*. We have tested 4-year-olds' sensitivity to the spacing among facial features under a range of task conditions: judgments of distinctiveness, matching facial identity in the absence of memory demands, and recognizing the identity of personally familiar faces. Sensitivity to the spacing among facial features appears to allow adults to recognize a face from a new point of view (Mondloch, Geldart, Maurer, & Le Grand, 2003). Our results predict, therefore, that 4-year-old children may perform poorly when asked to recognize a face from a novel point of view and under other conditions that make featural cues less reliable (e.g., under poor lighting conditions). Under many conditions, however, 4-year-old children will recognize faces with great accuracy; they are very sensitive both to differences among faces in facial contour and in the appearance of individual features (e.g., the eyes and mouth; Freire & Lee, 2001; Mondloch et al., 2006)—other important cues to facial identity. The contribution of sensitivity to other cues known to influence adults' ability to recognize faces (e.g., surface cues such as color and texture; Lee & Perrett, 1997; Russell, Biederman, Nederhouser, & Sinha, 2007) should also be investigated in future studies.

References

Barton, J. J. S., Zhao, J. & Keenan, J. P. (2003). Perception of global facial geometry in the inversion effect and prosopagnosia. *Neuropsychologica*, *4*, 1703–1711.

Bruce, V., Campbell, R. N., Doherty-Sneddon, G., Import, A., Langton, S., McAuley, S., et al. (2000). Testing face processing skills in children. *British Journal of Developmental Psychology*, *18*, 319–333.

Bruce, V., & Young, A. W. (1998). *In the eye of the beholder: The science of face perception*. Oxford, UK: Oxford University Press.

Campbell, R., & Tuck, M. (1995). Recognition of parts of famous-face photographs by children: An experimental note. *Perception*, *24*, 451–456.

Campbell, R., Walker, J., & Baron-Cohen, S. (1995). The development of differential use of inner and outer face features in familiar face identification. *Journal of Experimental Child Psychology*, *59*, 196–210.

Carver, L., Dawson, G., Panagiotides, H., Meltzoff, A. N., McPartland, J., Gray, J. et al. (2003). Age-related differences in neural correlates of face recognition during the toddler and preschool years. *Developmental Psychobiology*, *42*, 148–159.

De Sonneville, L. M. J., Verschoor, C. A., Njiokiktjien, C., Op het Veld, V., Toorenaar, N., & Vranken, M. (2002). Facial identity and facial emotions: Speed, accuracy, and processing strategies in children and adults. *Journal of Clinical and Experimental Neuropsychology*, *24*, 200–213.

Farkas, L. G. (1994). *Anthropometry of the head and face in medicine* (2nd ed.). New York: Elsevier.

Freire, A., & Lee, K. (2001). Face recognition in 4- to 7-year-olds: Processing of configural, featural, and paraphernalia information. *Journal of Experimental Child Psychology*, *80*, 347–371.

Ge, L., Luo, J., Nishimura, M., & Lee, K. (2003). The lasting impression of chairman Mao: Hyperfidelity of familiar-face memory. *Perception*, *32*, 601–614.

Gilchrist, A., & McKone, E. (2003). Early maturity of face processing in children: local and relational distinctiveness effects in 7-year-olds. *Visual Cognition*, *10*, 769–793.

Goffaux, V., & Rossion, B. (2007). Face inversion disproportionately impairs the perception of vertical but not horizontal relations between features. *Journal of Experimental Psychology: Human Perception and Performance*, *33*, 995–1002.

Hayden, A., Bhatt, R. S., Reed, A., Corbly, C. R., & Joseph, J. E. (2007). The development of expert face processing: Are infants sensitive to normal differences in second-order relational information? *Journal of Experimental Child Psychology*, *97*, 85–98.

Hole, G. (1994). Configural factors in the perception of unfamiliar faces. *Perception, 23*, 65–74.

Lee, K. J., & Perrett, D. (1997). Presentation-time measures of the effects of manipulations in colour space on discrimination of famous faces. *Perception, 26*, 733–752.

Le Grand, R., Mondloch, C., Maurer, D., & Brent, H. P. (2001). Early visual experience and face processing. *Nature, 410*, 890 (Correction: 2001, *412*, 786).

Le Grand, R., Mondloch, C., Maurer, D., & Brent, H. (2003). Expert face processing requires visual input to the right hemisphere during infancy. *Nature Neuroscience, 6*, 1108–1112 (Erratum: 2003, *6*, 1329).

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Science, 6*, 255–260.

McKone, E., & Boyer, B. L. (2006). Sensitivity of 4-year-olds to featural and second-order relational changes in face distinctiveness. *Journal of Experimental Child Psychology, 94*, 134–162.

Mondloch, C. J., Dobson, K. S., Parsons, J., & Maurer, D. (2004). Why 8-year-olds can't tell the difference between Steve Martin and Paul Newman: Factors contributing to the slow development of sensitivity to the spacing of facial features. *Journal of Experimental Child Psychology, 89*, 159–181.

Mondloch, C. J., Geldart, S., Maurer, D., & Le Grand, R. (2003). Developmental changes in face processing skills. *Journal of Experimental Child Psychology, 86*, 67–84.

Mondloch, C. J., Leis, A., & Maurer, D. (2006). Recognizing the face of Johnny, Suzy, and me: Insensitivity to the spacing among features at 4 years of age. *Child Development, 77*, 234–243.

Mondloch, C. J., Le Grand, R., & Maurer, D. (2002). Configural face processing develops more slowly than featural face processing. *Perception, 31*, 553–566.

Mondloch, C. J., Le Grand, R., & Maurer, D. (2003). Early visual experience is necessary for the development of some—but not all—aspects of face processing. In O. Pascalis & A. Slater (Eds.), *The development of face processing in infancy and early childhood* (pp. 99–117). New York: Nova.

Mondloch, C. J., & Maurer, D. (2003). [Four-year-old children's sensitivity to three manipulations of grotesqueness in faces: featural changes, thatcherization, and spatial alterations]. Unpublished raw data.

Pellicano, E., & Rhodes, G. (2003). Holistic processing of faces in preschool children and adults. *Psychological Science, 14*, 618–622.

Pellicano, E., Rhodes, G., & Peters, M. (2006). Are preschoolers sensitive to configural information in faces? *Developmental Science, 9*, 270–277.

Russell, R., Biederman, I., Nederhouser, M., & Sinha, P. (2007). The utility of surface reflectance for the recognition of upright and inverted faces. *Vision Research, 47*, 157–165.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology, Human Experimental Psychology, 46a*, 225–245.

Tanaka, J.W., & Farah, M. J. (2003). Holistic face recognition. In M. Peterson & G. Rhodes (Eds.), *Analytic and holistic processes in the perception of faces, objects and scenes* (pp. 53–91). New York: Oxford University Press.

Tanaka, J. W., Kay, J. B., Grinnell, E., Stansfield, B., & Szechter, T. (1998). Face recognition in young children: When the whole is greater than the sum of its parts. *Visual Cognition, 5*, 479–496.

Tanaka, J. W., & Sengo, J. A. (1997). Features and their configuration in face recognition. *Memory & Cognition, 25*, 583–592.

Child Development, September/October 2008, Volume 79, Number 5, Pages 1524 – 1537

Limitations to Teaching Children 2 + 2 = 4: Typical Arithmetic Problems Can Hinder Learning of Mathematical Equivalence

Nicole M. McNeil
University of Notre Dame

Do typical arithmetic problems hinder learning of mathematical equivalence? Second and third graders (7 – 9 years old; $N = 80$) received lessons on mathematical equivalence either with or without typical arithmetic problems (e.g., $15 + 13 = 28$ vs. $28 = 28$, respectively). Children then solved math equivalence problems (e.g., $3 + 9 + 5 = 6 + \_\_$), switched lesson conditions, and solved math equivalence problems again. Correct solutions were less common following instruction with typical arithmetic problems. In a supplemental experiment, fifth graders (10 – 11 years old; $N = 19$) gave fewer correct solutions after a brief intervention on mathematical equivalence that included typical arithmetic problems. Results suggest that learning is hindered when lessons activate inappropriate existing knowledge.

Some of the most well-known and widely studied phenomena in developmental psychology involve children's failures to learn in seemingly straightforward situations. For example, 3-year-old children continue to sort cards on the basis of one dimension (e.g., color), despite being told repeatedly that they should start sorting according to another dimension (e.g., shape; Zelazo, Frye, & Rapus, 1996), and 9-year-old children continue to hold their initial, incorrect hypotheses about factors that influence the speed of a racecar (e.g., presence of a muffler), even after observing the outcomes of several experiments that yield results to the contrary (Schauble, 1990). Why do children sometimes fail to learn in seemingly straightforward situations, despite substantial instruction? This question has broad theoretical and practical implications. Although it is tempting to attribute children's failures to general conceptual limitations in childhood, several researchers have noted the importance of domain knowledge and basic learning processes in children's failures (see Siegler, 2000).

Here, we focus on children's failures to learn a seemingly straightforward topic in mathematics—mathematical equivalence. *Mathematical equivalence* is the relation between two quantities that are interchangeable. It can be expressed symbolically in the form of a math sentence (e.g., $3 + 4 = 7$) or equation (e.g., $x + 1 = 5$) containing two mathematical expressions separated by the equal sign. The symbolic form implies that the expression on the left side of the equal sign can be replaced by the expression on the right side and vice versa. Although there are nuances that distinguish equivalence from other relations such as equality and identity, these nuances are beyond the scope of the present study. The term *mathematical equivalence* will be used throughout this article because it is the most inclusive and flexible (Kieran, 1981). The term will be used to refer specifically to mathematical equivalence in symbolic form. The key issue here is that we want children to be able to interpret math sentences and equations that contain the equal sign *relationally* as statements of mathematical equivalence because a relational understanding of the equal sign is important for understanding algebra (Carpenter, Franke, & Levi, 2003; Kieran, 1989; Knuth, Stephens, McNeil, & Alibali, 2006).

It is well documented that elementary school children (aged 7 – 11 years) in the United States have substantial difficulty learning to interpret the equal sign as a relational symbol of mathematical equivalence (Baroody & Ginsburg, 1983; Behr, Erlwanger, & Nichols, 1980; Kieran, 1981). Consider, for example, a task in which children are asked to check the "correctness" of math sentences written by a girl or boy who "attends another school." Most children mark sentences such as $10 = 6 + 4$ as "incorrect" and change them to $6 + 4 = 10$, $4 + 6 = 10$, or even $10 +$

This research was supported by an APA Dissertation Award. Portions of these data were presented at the 2004 meeting of the Cognitive Science Society. Preparation of this article was supported in part by Grant R305B070297 from the U.S. Department of Education, Institute of Education Sciences. Thanks to Martha Alibali, Julia Evans, Colleen Moore, Jenny Saffran, Mark Seidenberg, and members of the Cognitive Development Research Group at the University of Wisconsin for their input. Thanks also to Heather Brletic-Shipley for her enthusiasm.

Correspondence concerning this article should be addressed to Nicole M. McNeil, Department of Psychology, University of Notre Dame, 118 Haggar Hall, Notre Dame, IN 46556. Electronic mail may be sent to nmcneil@nd.edu.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7905-0020

6 = 4 (Cobb, 1987; see also Baroody & Ginsburg, 1983; Behr et al., 1980; Rittle-Johnson & Alibali, 1999). Similar misconceptions are also apparent when children are asked to solve addition problems that have operations on both sides of the equal sign (e.g., 3 + 4 + 5 = 3 + __). These problems have been called "mathematical equivalence problems" in prior work (Perry, 1991; Perry, Church, & Goldin-Meadow, 1988). In the absence of instruction, approximately 82% of children (aged 7 – 11 years) solve mathematical equivalence problems incorrectly (McNeil, 2005), and even after receiving instruction, many children continue to have difficulty (Alibali, 1999; Perry, 1991; Rittle-Johnson & Alibali, 1999).

Why do children have trouble learning mathematical equivalence, and how can we help them overcome their difficulties? Contrary to long-established theories of children's (mis)understanding of mathematical equivalence (e.g., Collis, 1974, as cited in Kieran, 1981; Piaget & Szeminska, 1941/1995), it does not appear as though children's difficulties can be attributed to domain general conceptual limitations in childhood. Indeed, children in this age range succeed when asked to identify which pair of numbers (e.g., 3 + 4, 7 + 1, 5 + 6) is equal to a given pair (e.g., 5 + 3), suggesting that they can identify an equivalence relation between two addend pairs (Rittle-Johnson & Alibali, 1999). Children in this age range also are relatively good at solving math equivalence problems when they are presented in a semi-symbolic form in which faces of dice are used in place of the Arabic numerals, and the answer can be chosen from one of four possible answers (Sherman, 2007). Moreover, a majority (~ 90%) of Chinese elementary school children solve math equivalence problems correctly, suggesting that age-related conceptual limitations are not the primary source of children's difficulties (Li, Ding, Capraro, & Capraro, 2008).

The emerging consensus is that children's difficulties with mathematical equivalence are caused, at least in part, by specific knowledge that children construct from their early experience with arithmetic in school (e.g., Baroody & Ginsburg, 1983; Carpenter et al., 2003; McNeil & Alibali, 2005b; Seo & Ginsburg, 2003). In the United States, elementary school mathematics is monopolized by arithmetic (Beaton et al., 1996), which is usually taught in a very procedural fashion with little or no reference to the equal sign or the concept of mathematical equivalence. Moreover, typical arithmetic problems are almost always presented with the operations to the left of the equal sign and the "answer blank" to the right (e.g., 3 + 4 = __; Seo & Ginsburg, 2003), a format that does not highlight the interchangeable nature of the two sides of the equation. This narrow experience with arithmetic leads children to construct overly narrow knowledge structures that do not generalize beyond typical arithmetic problems (McNeil & Alibali, 2005b). Reliance on these knowledge structures, in turn, contributes to difficulties with higher order problems, including math equivalence problems. Note that the claim is not necessarily that experience with arithmetic per se contributes to the difficulties, but rather that experience with arithmetic *as typically taught in U.S. schools* contributes to the difficulties.

Prior work has identified at least three patterns that children learn from their experience with arithmetic in school. These patterns have been called *operational patterns* (McNeil & Alibali, 2005b). First, children learn a perceptual pattern related to the format of math problems, namely, the "operations = answer" format in which the operations are to the left of the equal sign and the answer blank is to the right (McNeil & Alibali, 2004; Seo & Ginsburg, 2003). Second, they learn a concept of the equal sign, namely, that it means "calculate the total" (Baroody & Ginsburg, 1983; Behr et al., 1980; Kieran, 1981; McNeil & Alibali, 2005a). Third, they learn a strategy for solving math problems, namely, "perform all the given operations on all the given numbers" (McNeil & Alibali, 2005b; Perry et al., 1988). With continued practice, children's representations of these operational patterns increase in strength. Subsequently, when these representations are activated in the face of math problems, they guide attention and influence how information is encoded, interpreted, and ultimately applied (cf. Bruner, 1957; McGilly & Siegler, 1990; McNeil & Alibali, 2004). Although this process likely facilitates fast and accurate performance on typical arithmetic problems, it may interfere with performance on problems that deviate from the typical form, such as math equivalence problems.

In accordance with this view, studies have shown that children do, indeed, misapply their knowledge of typical arithmetic problems to math equivalence problems. For example, when asked to reconstruct the problem 3 + 4 + 5 = 3 + __ after viewing it briefly, children rely on their knowledge of the typical "operations = answer" problem format and write "3 + 4 + 5 + 3 = __" (McNeil & Alibali, 2004). When asked to define the equal sign in the problem, children provide answers such as "it means what all the numbers add up to" (McNeil & Alibali, 2005a). When asked to solve the problem, children add up all the numbers and put "15" in the blank. This "add-all" strategy is the most common strategy used by children who have not received instruction on mathematical equivalence (McNeil, 2005; Perry et al., 1988), and use of this strategy increases as the amount of experience with traditional arithmetic increases from first to third grade (McNeil, 2007). Moreover, a recent study has even suggested that children's reliance on their knowledge of these operational patterns may hinder their ability to benefit from a brief one-on-one intervention on math equivalence problems (McNeil & Alibali, 2005b).

Although much of the available evidence supports the hypothesis that children's narrow knowledge of arithmetic hinders learning of mathematical equivalence, experimental evidence is lacking. One exception is an experiment by McNeil and Alibali (2005b, Experiment 2) in which undergraduates' knowledge of arithmetic was activated or was not activated in a control condition. Undergraduates whose knowledge of arithmetic was activated were less likely to perform well on math equivalence problems. This finding is important from a theoretical standpoint because it shows that the activation of arithmetic knowledge can interfere with performance on math equivalence problems. However, the practical significance is limited. The finding indicates that activating *undergraduates'* knowledge of arithmetic hinders *performance* on math equivalence problems in a *laboratory* setting, but we primarily want to know how activating *children's* knowledge of arithmetic affects their *learning* in *classroom* settings. If similar processes affect children's learning of mathematical equivalence, then aspects of the environment that activate and reinforce children's overly narrow knowledge of arithmetic may make it more difficult for children to learn mathematical equivalence.

If the activation of children's knowledge of typical arithmetic problems does indeed hinder children's learning of mathematical equivalence, then teachers may want to avoid lesson contexts that contain typical arithmetic problems when trying to improve children's understanding of mathematical equivalence. Different contexts activate different representations (e.g., Barsalou, 1982), and the context of a typical arithmetic problem activates children's overly narrow representations of the operational patterns (McNeil & Alibali, 2005a; McNeil et al., 2006). Granted, elementary school children may activate these representations to some degree any time they see a math sentence or equation containing the equal sign; however, the likelihood and magnitude of that activation will be greatest in the context of a typical arithmetic problem because the problem maps directly onto the operational patterns (McNeil & Alibali, 2004; McNeil et al., 2006). Once children's knowledge of the operational patterns is activated, children may be less open to information that deviates from these patterns. This perspective suggests a somewhat counterintuitive method of remediation. That is, instead of teaching mathematical equivalence by presenting familiar arithmetic problems such as "2 + 2 = 4" and teaching children that "two plus two is the same amount as four," it might be better to teach mathematical equivalence in a context that does not contain typical arithmetic problems (e.g., present "4 = 4" and teach "four is the same amount as four"). According to this perspective, manipulations of the environment that are designed to activate children's knowledge of typical arithmetic problems should hinder learning of mathematical equivalence.

Although this prediction follows logically from the account that has been advanced thus far, there are also reasons to recommend the opposite. Indeed, intuition suggests that it might be easier for children to learn a new concept in the context of familiar content. Consistent with this view, many educators in the United States hold a review-based instruction philosophy, which assumes that previously learned material provides a supportive scaffold for new learning. This philosophy is evident in math classrooms in the United States where educators often implement a spiral mathematics curriculum in which old information is reintroduced year after year (see Jensen, 1990, for harms associated with the spiral mathematics curriculum). Similarly, several lines of research suggest that the most effective math and science instructions are the ones that "bridge" from children's well-established knowledge to new material (e.g., Bryce & MacMillan, 2005; Clement, Brown, & Zeitsman, 1989; Koedinger & Nathan, 2004; Nathan, Stephens, Masarik, Alibali, & Koedinger, 2002). Both the review-based philosophy and the bridging approach suggest that it may be beneficial for teachers to design lessons that use children's well-established knowledge (e.g., typical arithmetic) as support when introducing a new concept (e.g., mathematical equivalence).

Another reason to expect the presence of typical arithmetic problems to facilitate learning of mathematical equivalence is that it may provide an opportunity for cognitive conflict. Several theories suggest that cognitive conflict promotes learning and cognitive development (e.g., Limon, 2001; Piaget, 1980; VanLehn, 1996). These theories suggest that children will benefit from lessons that purposefully activate their incorrect, operational view of the equal sign and then contradict it. Previous research has shown that typical arithmetic problems such as "3 + 4 = __" activate the operational view (McNeil & Alibali, 2005a; McNeil et al., 2006). Thus, when an educator or parent teaches the relational meaning of the equal sign in the context of a typical arithmetic problem, it may push children into a state of cognitive conflict, which in turn may prompt them to reevaluate their operational view (cf. Limon, 2001).

The goal of the present study was to examine whether the presence of typical arithmetic problems undermines or enhances children's learning from lessons on mathematical equivalence. Children received an intervention that was designed to teach that the mathematical expression on the right side of the equal sign has to be the same amount as (equal to) the mathematical expression on the left side of the equal sign. Elementary school children understand what it means for two amounts to be equal (Rittle-Johnson & Alibali, 1999), so there was potential for them to benefit from this intervention. The intervention was given either in an *arithmetic condition* with typical arithmetic problems or in a *control condition* without typical arithmetic problems. After the intervention, children solved math equivalence problems. If activating children's knowledge of typical arithmetic problems interferes with learning of mathematical equivalence (as predicted), then children should benefit less from the intervention in the arithmetic condition than from the intervention in the control condition. That is, children should solve fewer math equivalence problems correctly after receiving lessons in the arithmetic versus control condition.

Experiment 1

Method

Participants

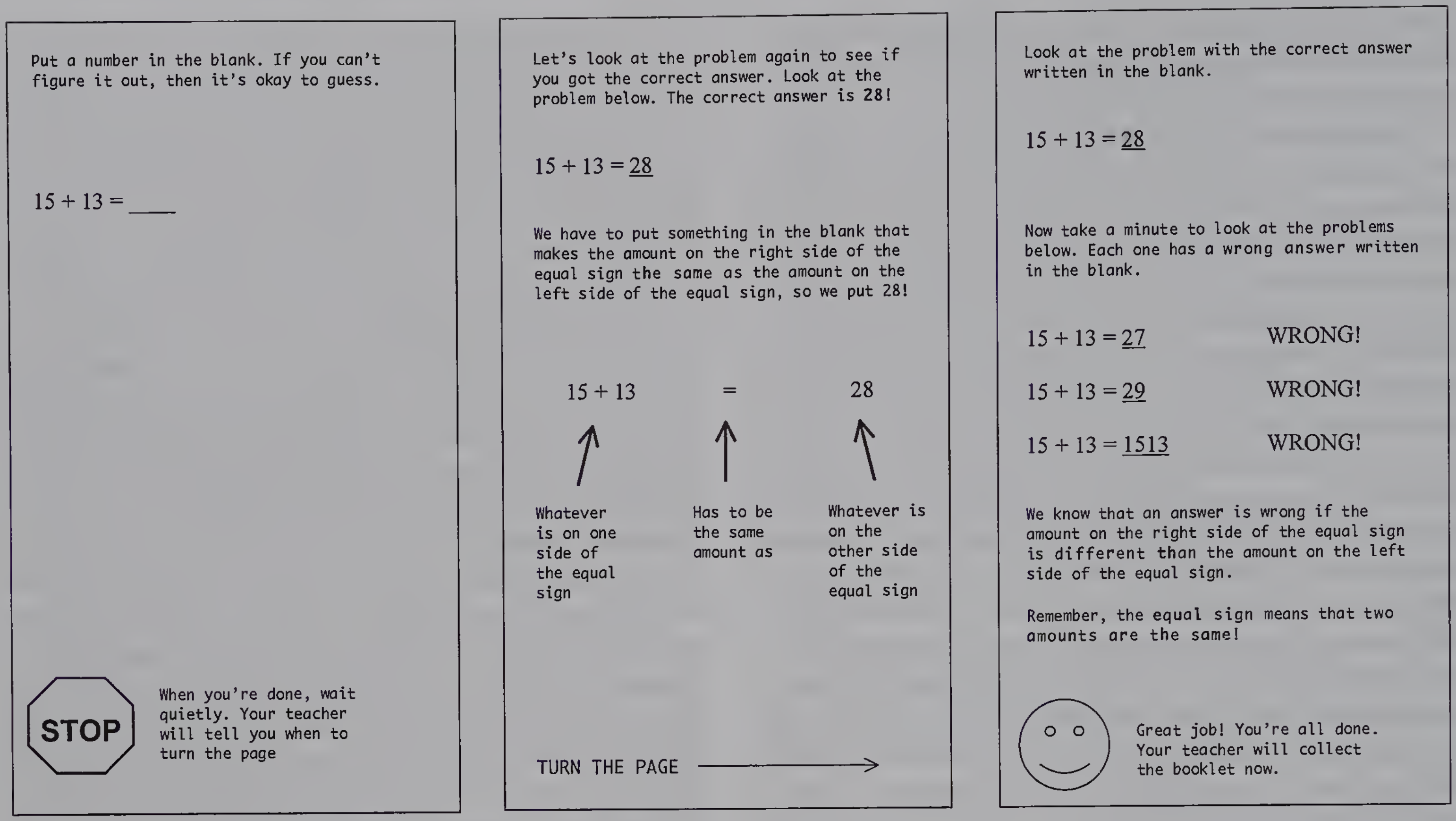
The study was conducted at a public school in a small town near Raleigh, North Carolina. The racial/ethnic makeup of the school was 24% African American, 1% Asian, 6% Hispanic, and 69% White. Approximately 37% of children received free or reduced-price lunch. Ninety-three second- and third-grade children participated. Eleven were excluded because they were absent on one or more days. Two were excluded because their performance was 3 *SD* away from the mean. The final sample contained 80 children (38 boys and 42 girls).

Materials

Scripted lessons on mathematical equivalence in symbolic form. Each daily lesson followed the same basic script, regardless of condition. The script had four main steps. First, teachers read the following instructions: "Today you are going to solve some problems. You may not talk to your neighbor, and you may not look at your neighbor's paper. Don't worry though! It's not a test, so you won't be graded. This is just to see what you think about the problems. If a problem seems too hard for you to solve, then it's okay for you to guess." Second, children were presented with a math problem and asked to solve it. Third, children were presented with a correctly solved version of the problem, and the concept of mathematical equivalence was used to justify the correct answer. Fourth, children were presented with three incorrectly solved versions of the problem, and the concept of mathematical equivalence was used to explain why the answers were wrong. Steps 2–4 were then repeated with a second problem. Children and teachers received booklets to follow along with each daily lesson. The teachers' booklets served as their script for the lessons. Figure 1 shows the script for one problem (15 + 13 = $\underline{28}$) as it was presented in the children's booklets. Given that children worked with different problems across conditions within the same classroom (see the Procedure section), the actual problems were not written in the teachers' booklets. Instead, all problems were replaced with the words "problem is here in students' booklet." With that exception, the teachers' booklets were identical to the children's.

Equation-solving test. Children were given unlimited time to solve 12 math equivalence problems, including 6 "right-blank" problems (e.g., 5 + 4 + 7 = 5 + __, 3 + 9 + 5 = 6 + __) and 6 "left-blank" problems (e.g., 9 + 5 + 7 = __ + 7, 6 + 4 + 8 = __ + 3). Problems of this type have been used extensively in previous research on children's understanding of mathematical equivalence (e.g., Alibali, 1999; McNeil & Alibali, 2005b; Perry et al., 1988). In order to solve these problems correctly without direct instruction on a correct procedure, children must understand that the equal sign is a symbol used to indicate that the mathematical expressions on either side of it are (or are set to be) equivalent. Children also must notice where the equal sign is in the problem and understand the rules for how numbers and operations can be manipulated in order to come up with a solution that makes both sides of the problem have the same value. Children's performance solving math equivalence problems correlates well with other measures designed to assess children's understanding of mathematical equivalence, such as defining the equal sign (Rittle-Johnson & Alibali, 1999) and encoding the structure of equations (McNeil & Alibali, 2004).

Prior to administering the equation-solving test, teachers read the following instructions: "Today we're going to see how well you can solve some problems. Don't worry! It's not a test, so you won't be



Put a number in the blank. If you can't figure it out, then it's okay to guess.

15 + 13 = ____

STOP

When you're done, wait quietly. Your teacher will tell you when to turn the page

Let's look at the problem again to see if you got the correct answer. Look at the problem below. The correct answer is 28!

15 + 13 = 28

We have to put something in the blank that makes the amount on the right side of the equal sign the same as the amount on the left side of the equal sign, so we put 28!

15 + 13 = 28

Whatever is on one side of the equal sign

Has to be the same amount as

Whatever is on the other side of the equal sign

TURN THE PAGE →

Look at the problem with the correct answer written in the blank.

15 + 13 = 28

Now take a minute to look at the problems below. Each one has a wrong answer written in the blank.

15 + 13 = 27 WRONG!

15 + 13 = 29 WRONG!

15 + 13 = 1513 WRONG!

We know that an answer is wrong if the amount on the right side of the equal sign is different than the amount on the left side of the equal sign.

Remember, the equal sign means that two amounts are the same!

Great job! You're all done. Your teacher will collect the booklet now.

Figure 1. Script for one problem as it appeared in the students' booklets. Each panel represents one page.

graded. I just want you to do your best. After you finish solving the problems, close your booklet and sit quietly until everyone has finished. If a problem seems too hard for you to solve, then don't worry, you can just give it your best guess."

Procedure

The study was conducted over a 2-week period. Children's teachers administered the scripted lessons and collected the measures in the classroom setting. Teachers first taught 2 days of scripted lessons on mathematical equivalence. Lessons were on Tuesday and Thursday, and they lasted approximately 15–20 min each day. Then, children completed the 12 math equivalence problems on the equation-solving test on Friday. The following week, children again received scripted lessons on a Tuesday and Thursday in whichever condition they had not already received (e.g., children who already received lessons in the arithmetic condition now received lessons in the nonarithmetic condition). Finally, children completed the 12 problems on the equation-solving test again on Friday.

All children received the same *spoken* lessons because lessons were scripted. The lessons taught that the amount on the right side of the equal sign has to be the same as (equal to) the amount on the left side of the equal sign. However, children differed according to what they were looking at during the lessons. Children were given booklets to follow along with the lessons, and children were randomly assigned to lesson conditions through the use of these booklets. In the *arithmetic condition*, booklets contained typical arithmetic problems (e.g., 15 + 13 = 28). In the *nonarithmetic condition*, booklets contained problems with the same solution designed to *not* activate knowledge of arithmetic (e.g., 28 = 28). Table 1 presents the problems used in each condition. As can be seen in the table, math equivalence problems were not presented in either lesson condition. Thus, children never received direct instruction on

Table 1
Problems Used in Each Lesson Condition

Arithmetic condition	Nonarithmetic condition
15 + 13 = 28	28 = 28
19–7 = 12	1 foot = 12 inches
6 + 5–8 = 3	●●● = 3
91 + 4 = 95	95 = 95

a procedure for solving math equivalence problems. Figure 2 presents a brief excerpt from each of the lesson conditions as it was presented in students' booklets.

The booklets enabled children *within the same classrooms* to be randomly assigned to conditions. This method of random assignment within classrooms is unique. Typically, teachers are randomly assigned to conditions, and children are nested within teachers. Randomly assigning children within classrooms offers several advantages, including higher power with fewer children and greater assurance that children across conditions are comparable. The development of this new methodology is an important contribution of the present study.

The final sample included 34 children who received lessons in the arithmetic condition first and 46 children who received lessons in the nonarithmetic condition first. No order effects were observed. That is, time point (first or second) did not have significant main effect on performance, and it did not interact with lesson condition to affect performance on equivalence problems (all p values $> .10$). Nonetheless, to ensure the reliability of the reported results, three additional analyses were performed in which 12 of the 46 children who received the nonarithmetic condition first were omitted randomly (to simulate equal n). Conclusions based on the "equal-n" analyses were the same as those reported here.

Coding

Children's performance on the math equivalence problems on the equation-solving test was coded using a system developed by Perry et al. (1988). Children's use of a correct strategy on a given problem was based on an objective criterion—the solution children wrote in the blank. As in past work, children were given credit for using a correct strategy as long as they were within ±1 of the correct answer. For example, on the problem $4 + 8 + 9 = 4 + \_\_$, children would be given credit for using a correct strategy if they put 17, 16, or 18 in the blank.

Results and Discussion

Performance on the equation-solving test was poor. Even after receiving several lessons on mathematical equivalence from their classroom teachers, children solved only 0.73 ($SD = 0.98$) math equivalence problems correctly (out of 12). This poor performance is consistent with prior work (e.g., Alibali, 1999; McNeil & Alibali, 2005b, Experiment 1; Perry et al., 1988), and it highlights just how difficult it is for children to learn and apply mathematical equivalence. Poor performance was characteristic of both second-grade children ($M = 0.69$, $SD = 0.52$) and third-grade children ($M = 0.74$, $SD = 0.83$).

A repeated-measures analysis of variance (ANOVA) was performed with lesson condition (arithmetic or nonarithmetic) as the independent variable and number correct on the equation-solving test (of 12) as the dependent measure. There was a significant main effect of lesson condition, $F(1, 79) = 5.40, p = .023, \eta_p^2 = .06$. As predicted, children solved fewer problems correctly after receiving lessons in the arithmetic condition ($M = 0.56$, $SD = 0.90$) than after receiving lessons in the nonarithmetic condition ($M = 0.89$, $SD = 1.03$). However, it is important to note that children's performance on the math equivalence problems was very poor overall.

Given the potential floor effects, a nonparametric analysis is preferred. Therefore, in accordance with previous studies (e.g., McNeil, 2007; McNeil & Alibali, 2005b), we reanalyzed the data in terms of whether children solved at least one math equivalence problem correctly. Across conditions, 65% of children solved at least one math equivalence

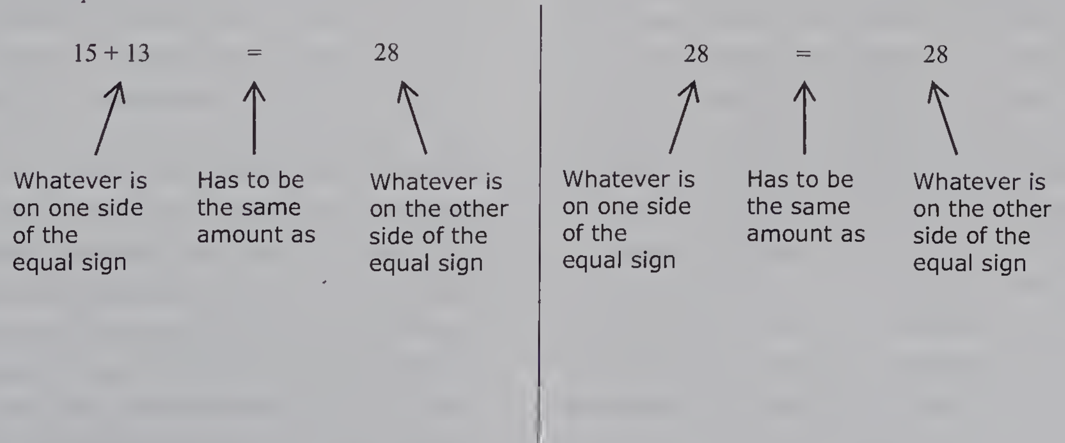

15 + 13
=
28
Whatever is on one side of the equal sign
Has to be the same amount as
Whatever is on the other side of the equal sign
28
=
28
Whatever is on one side of the equal sign
Has to be the same amount as
Whatever is on the other side of the equal sign

Figure 2. Brief excerpt from the students' lesson booklets in the arithmetic context (left panel) and nonarithmetic context (right panel).

problem correctly. Previous studies have shown a much lower percentage (around 10%–20%) for children in second and third grades who have not received lessons on mathematical equivalence; therefore, this result suggests that the lessons did improve understanding of the symbolic form of mathematical equivalence, albeit weakly. Note that it is not a trivial matter for children to solve one math equivalence problem correctly. Children were free to write any number in the blank. Thus, the probability that children were able to generate the correct answer by guessing is low. Consistent with McNeil (2007), a greater percentage of second-grade children (74%) than third-grade children (60%) solved at least one math equivalence problem correctly.

We examined the effect of lesson condition on performance using a nonparametric test appropriate for repeated measures data. The McNemar test was used to compare children's likelihood of solving at least one math equivalence problem correctly after receiving lessons in the arithmetic condition versus the nonarithmetic condition. Results were consistent with the ANOVA. Of the 46 children who did not solve at least one math equivalence problem correctly after receiving lessons in the arithmetic condition, 18 (39%) solved at least one math equivalence problem correctly after receiving lessons in the nonarithmetic condition; however, only 4 (12%) of the 32 children who did not solve at least one math equivalence problem correctly after receiving lessons in the nonarithmetic condition solved at least one math equivalence problem correctly after receiving lessons in the arithmetic condition, $\chi^2(1, N = 80) = 8.91$, $p = .003$.

In summary, the presence of typical arithmetic problems hindered children's learning from lessons on mathematical equivalence. Given the same verbal lessons, children solved fewer math equivalence problems correctly on average after receiving lessons in the arithmetic condition than after receiving lessons in the nonarithmetic condition, and a smaller percentage of children solved at least one math equivalence problem correctly after receiving lessons in the arithmetic versus nonarithmetic condition. We suspect that the negative effects of the arithmetic condition are due to the overly narrow knowledge structures activated by typical arithmetic problems, specifically knowledge of the "operations = answer" problem format, the "calculate the total" interpretation of the equal sign, and the "perform all the given operations" problem-solving strategy. When these representations are activated, they have the potential to hinder learning of information that does not correspond to the typical form.

Although these results support the hypothesis that the activation of knowledge of typical arithmetic problems hinders learning of mathematical equivalence, there are at least two reasons to be cautious when interpreting the results. First, the experiment compared lessons presented in an arithmetic condition to lessons presented in a nonarithmetic condition. The nonarithmetic condition exposed children to nonstandard equation types (e.g., 28 = 28), which are rarely presented in elementary mathematics instruction (Seo & Ginsburg, 2003). Thus, although we found evidence in support of the hypothesis that the arithmetic condition hinders learning of mathematical equivalence, another plausible interpretation of the results is that the nonarithmetic condition facilitates learning. To rule out this alternative hypothesis, the arithmetic condition would need to be compared to a context-free, control condition. Thus, we tested this comparison in a supplemental experiment (Experiment 2).

The second reason to be cautious is that performance on the math equivalence problems was very poor overall. Children in this age range perform well on typical arithmetic problems in which the goal is to determine the numerical value of several numbers and operations presented on the left side of the equal sign (e.g., 3 + 4 + 5 + 3 = __; McNeil & Alibali, 2004). They also perform well when asked to choose which addend pair (e.g., 1 + 6, 3 + 5, 2 + 4) is equal to a given addend pair (e.g., 4 + 4; Rittle-Johnson & Alibali, 1999). However, they struggle to solve math equivalence problems. The present experiment was no exception. Although the difference between the two lesson conditions in terms of use of a correct strategy was statistically significant and did not depend on the method of analysis, neither condition was very effective in increasing children's use of correct strategies on math equivalence problems. This limitation may be minor from a theoretical standpoint because results supported the hypothesis. Nonetheless, from a practical standpoint, it would be more convincing if the effect could be replicated with a group of children who could reap more benefits from the intervention.

Unfortunately, finding such children is not a trivial task. It is a delicate balance. We need to find children for whom the operational way of thinking about equations is dominant prior to instruction but who, at the same time, are likely to benefit from a brief intervention on mathematical equivalence. One possible solution is to screen older, more experienced children. Children who have more experience with a variety of mathematics topics (e.g., equivalent fractions, magnitude comparison problems) are less entrenched in the operational way of thinking

(McNeil, 2007). They have the potential to exhibit both operational and relational ideas about equations depending on the context (McNeil & Alibali, 2005a; McNeil et al., 2006), so they should be more open to a brief intervention on mathematical equivalence. However, there may also be drawbacks to working with older children. Namely, it may be difficult to find older, more experienced children for whom the operational way of thinking about equations is dominant. Although older children continue to exhibit operational ideas about equations that hinder understanding of higher order math topics in middle school and beyond (Kieran, 1981; Knuth et al., 2006; McNeil & Alibali, 2005b; McNeil et al., 2006), they also exhibit relational ideas about equations in certain contexts (McNeil & Alibali, 2005a; McNeil et al., 2006) and solve math equivalence problems correctly (McNeil, 2007). It is crucial to work with children who still maintain the operational way of thinking about equations as their dominant view, so we can examine which conditions encourage (or discourage) children from moving beyond that view. At the same time, we need to work with children who have more experience with a broader array of mathematics topics so we can increase the likelihood that they will be more open to an intervention than were the children in Experiment 1. Thus, we screened for such children in Experiment 2.

Experiment 2

Experiment 2 was designed to address two specific goals. First, we sought to compare the arithmetic condition to a control condition that did not include nonstandard equations. Second, we sought to test a group of older children who still exhibited the operational way of thinking about equations as their dominant view. We expected these children to be more likely than the second- and third-grade children to use correct strategies on math equivalence problems after a brief intervention.

Method

Participants

The study was conducted at a suburban public school in Connecticut. The racial/ethnic makeup of the school was 3% African American, < 1% Asian, 2% Hispanic, and 94% White. Approximately 10% of children received free or reduced-price lunch. The children were selected from 57 fifth-grade children who received consent to participate. Of those children, 19 (11 boys and 8 girls) met criteria for participation because they interpreted the equal sign as an arithmetic operation on the pretest (see the Coding section for details). Of the 38 children who did not meet the criteria, 31 had other incorrect interpretations of the equal sign, and 7 had a relational interpretation of the equal sign. The average reported math grades of children who met criteria for inclusion did not differ significantly from the average reported math grades of children who did not meet criteria, $F(1, 55) = 0.80$, $p = .37$.

Measures

Pretest measure of equal sign understanding. The measure of equal sign understanding was modeled after measures used in previous work (e.g., Knuth et al., 2006; McNeil & Alibali, 2005a; Rittle-Johnson & Alibali, 1999). There was an arrow pointing to the equal sign followed by two questions: (a) "The arrow above points to a symbol. What is the name of the symbol?" and (b) "What does the symbol mean?" This measure has good convergent validity (e.g., McNeil & Alibali, 2005a) and good concurrent validity (e.g., Knuth et al., 2006).

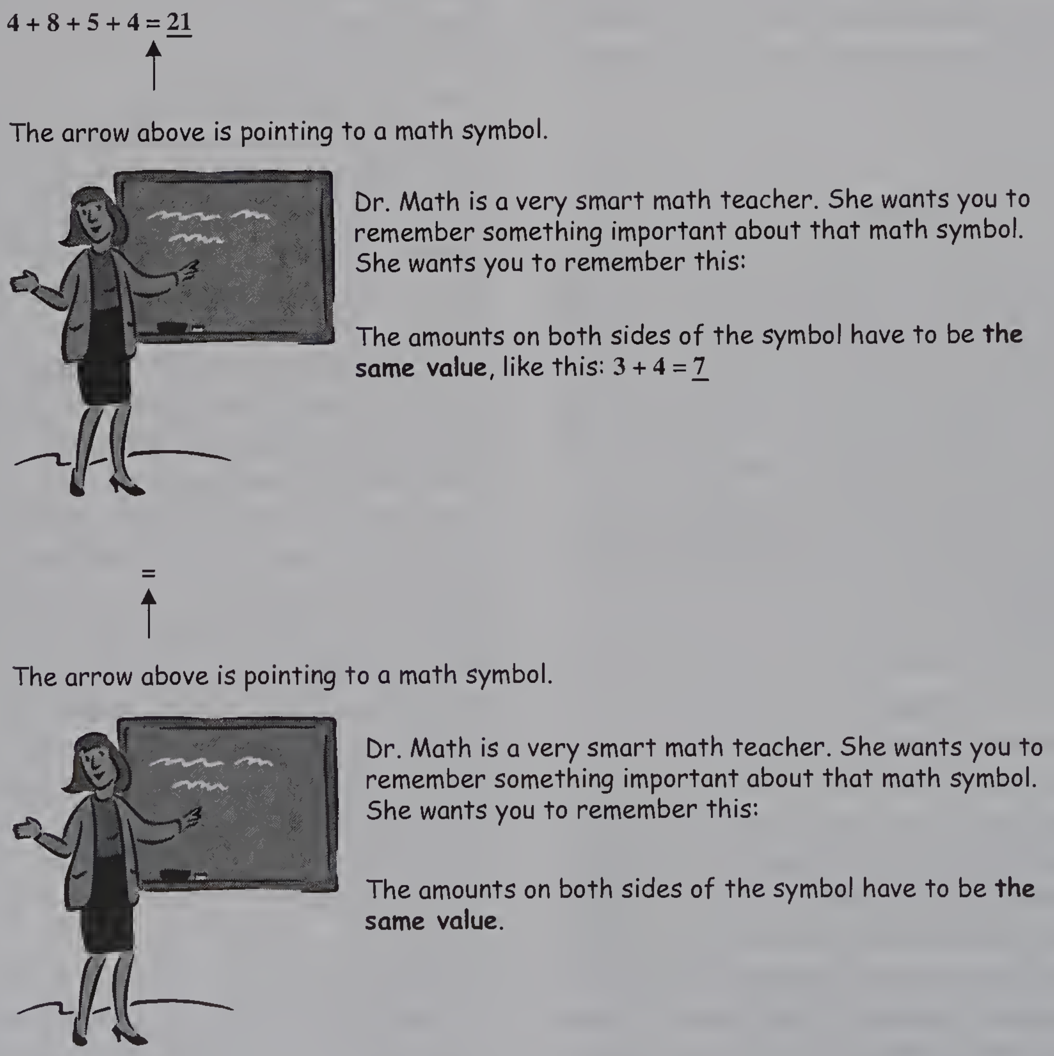
Equation-solving posttest. Children were given unlimited time to solve 4 right-blank math equivalence problems: $5 + 4 + 8 = 5 + \_\_$, $4 + 8 + 9 = 4 + \_\_$, $3 + 9 + 7 = 6 + \_\_$, and $9 + 4 + 6 = 3 + \_\_$. We used 4 problems instead of 12 in favor of efficiency because past work has shown similar performance regardless of whether children solve 3 (e.g., McNeil & Alibali, 2000), 4 (e.g., Rittle-Johnson & Alibali, 1999), or greater than 4 math equivalence problems (e.g., Experiment 1; Alibali, 1999; McNeil & Alibali, 2004; Perry, 1991; Perry et al., 1988). We used only right-blank problems because several studies have used only right-blank problems (e.g., McNeil & Alibali, 2000; Rittle-Johnson & Alibali, 1999), and children in prior studies have been equally likely to use incorrect strategies on right- and left-blank problems (e.g., McNeil & Alibali, 2004). Indeed, children's use of a correct problem-solving strategy does not seem to depend on (a) the location of the missing addend, (b) the symbol used in place of the missing addend (__ vs. □), or (c) the number of addends on the left side of the problem (see Carpenter et al., 2003; Ding et al., 2007). Children are just as likely to use an incorrect strategy on the problem $5 + 4 + 8 = 5 + \_\_$ as they are on the problem $\square + 8 = 12 + 5$.

Procedure

The experiment was held in school during regular school hours. Children were given the pretest measure, the intervention, and the posttest measure as paper-and-pencil worksheets in a large-group setting. First, children completed the pretest measure of equal sign understanding. Then, they were randomly assigned to one of two brief interventions on the meaning of the equal sign. Figure 3 displays the two interventions as children saw them. As shown in the figure, the interventions taught the same information about the equal sign—the amounts on both sides have to be the same value. Both interventions contained this conceptual information about the meaning of the equal sign. The factor that was manipulated between the interventions was the context in which the equal sign was presented. In the *arithmetic context*, the equal sign was presented in the context of a typical arithmetic problem (e.g., $3 + 4 = \underline{7}$). In the *alone context*, the equal sign was presented alone, without any associated math context (e.g., =). The alone context was used instead of the nonstandard equation types that were used in Experiment 1 (e.g., $28 = \underline{28}$) to ensure that any differences between conditions could be attributed to the negative effect of the typical arithmetic problems and not to the beneficial effect of nonstandard equation types.

Coding

Equal sign understanding. Children's responses on the pretest measure of equal sign understanding were coded based on the system used by McNeil and Alibali (2005b). This system differs from previous coding systems (e.g., McNeil & Alibali, 2005a; Rittle-Johnson & Alibali, 1999) because it distinguishes between different types of incorrect definitions of the equal sign. Specifically, it distinguishes children who mention arithmetic operations and/or arithmetic symbols in their definitions (e.g., "it means you need to add, subtract, multiply, or divide") from children who do not (e.g., "it means what the question is"). We know that children who mention arithmetic operations and/or arithmetic symbols in their definitions are more deeply entrenched in the operational view of equations than are children who provide other incorrect definitions of the equal sign (McNeil & Alibali, 2005b). Because our goal in the present experiment was to work with older children who are still entrenched in the operational view, we specifically screened for children who mentioned arithmetic operations and/or arithmetic symbols in their



$4 + 8 + 5 + 4 = \underline{21}$

The arrow above is pointing to a math symbol.

Dr. Math is a very smart math teacher. She wants you to remember something important about that math symbol. She wants you to remember this:

The amounts on both sides of the symbol have to be **the same value**, like this: $3 + 4 = \underline{7}$

=

The arrow above is pointing to a math symbol.

Dr. Math is a very smart math teacher. She wants you to remember something important about that math symbol. She wants you to remember this:

The amounts on both sides of the symbol have to be **the same value**.

Figure 3. The equal sign intervention in the arithmetic context (top panel) and alone context (bottom panel).

definitions of the equal sign. Reliability was established by having a second coder evaluate the interpretations of a randomly selected 20% sample ($N = 11$). Agreement between coders was 100%.

Equation-solving performance. Coding of equation-solving performance was identical to that in Experiment 1.

Results and Discussion

The fifth-grade children's performance on the equation-solving test was relatively good compared to the performance of second- and third-grade children in Experiment 1. After receiving the brief intervention on the meaning of the equal sign, children solved 2.10 ($SD = 1.97$) math equivalence problems correctly (out of 4). This finding is consistent with recent work showing that performance on math equivalence problems is better after age 9 (McNeil, 2007). An ANOVA was performed with intervention context as the independent variable and number correct on the equation-solving measure (out of 4) as the dependent measure. Children who received the intervention in the typical arithmetic context solved fewer problems correctly than did children who received the intervention in the alone context ($M = 1.00$, $SD = 0.50$ vs. $M = 3.10$, $SD = 0.42$), $F(1, 17) = 7.26$, $p = .01$, $\eta_p^2 = .30$. Thus, there was a large effect of intervention context on performance (as indicated by the effect size). Another way to analyze the data is to use a nonparametric test to examine whether condition affects whether or not children ever solved at least one math equivalence problem correctly. Only 3 of 9 children (33%) solved at least one math equivalence problem correctly after receiving the intervention in the arithmetic context, whereas 8 of 10 children (80%) solved at least one math equivalence problem correctly after receiving the intervention in the alone context, $\chi^2(1, N = 19) = 4.23$, $p = .04$.

Taken together with the results of Experiment 1, these findings support the hypothesis that the activation of children's knowledge of typical arithmetic problems hinders their learning of mathematical equivalence. When a typical arithmetic problem is present in the environment, it activates children's representations of the operational patterns. The activation of these representations, in turn, hinders learning of information that deviates from the typical arithmetic form. Findings suggest that learning of mathematical equivalence is hindered when overly narrow knowledge of typical arithmetic is activated. The potential processes involved in these effects and the implications are discussed in the next section.

General Discussion

Over 20 years of research suggests that children's difficulties with mathematical equivalence are caused, at least in part, by specific knowledge that children construct from their narrow experience with typical arithmetic problems in school (e.g., Baroody & Ginsburg, 1983; Carpenter et al., 2003; McNeil, 2007). According to this view, children should have more difficulty learning mathematical equivalence when their narrow knowledge is activated than when it is not activated. The present study was the first to test this hypothesis experimentally. In Experiment 1, second- and third-grade children were less likely to solve math equivalence problems (e.g., $3 + 9 + 5 = 6 + \_\_$) correctly after receiving lessons in the presence of typical arithmetic problems than after receiving lessons in the presence of other math problems. In Experiment 2, fifth-grade children who still interpreted the equal sign as an arithmetic operation solved fewer math equivalence problems correctly after receiving an intervention in the presence of typical arithmetic problems than after receiving an intervention in the presence of the equal sign alone. Both experiments support the hypothesis that the presence of typical arithmetic problems hinders learning of mathematical equivalence.

Given the theoretical and practical implications of these findings, it is important to understand the processes that may account for the observed effects. The explanation that has been advanced throughout this article is that the presence of typical arithmetic problems hinders learning because typical arithmetic problems activate overly narrow representations that do not generalize broadly. Specifically, typical arithmetic problems are thought to activate children's representations of (a) the "operations = answer" problem format, (b) the "calculate the total" interpretation of the equal sign, and (c) the "perform all the given operations" problem-solving strategy (McNeil & Alibali, 2005b). When these representations are activated, they have the potential to influence how information is encoded, interpreted, and used (cf. Bruner, 1957; McGilly & Siegler, 1990); therefore, information that does not map onto the patterns may not be learned as well as it could be in the absence of those activated representations. Although the results of the present study support this account, they do not rule out the possibility that other processes may have played a role in the observed effects.

For example, in addition to activating children's representations of specific patterns that interfere with learning (e.g., "operations = answer" problem format), the presence of typical arithmetic problems may trigger a more general cognitive bias, or mindset, that leads children to be inattentive to instructional input. Specifically, typical arithmetic problems may promote *mindlessness*, which has been defined as being committed to a "single, rigid perspective and . . . oblivious to alternative ways of knowing" (Langer, 2000). Mindlessness is a general state of mind that reduces individuals' sensitivity to the environment. Instead of noticing novel aspects of the environment and actively constructing strategies that apply to the current situation, mindless problem solvers are content to rely on the strategies they have used many times in the past. In the current study, the presence of typical arithmetic problems may have promoted mindlessness because children have seen and solved typical arithmetic problems many times in the past. If this is the case, then children may have been less attentive to the lessons in the arithmetic condition than in the control condition in both experiments.

It is also possible that some other systematic difference between the arithmetic and the control conditions (other than the presence of typical arithmetic problems per se) contributed to the observed effects. In Experiment 1, the control condition included nonstandard math sentences such as 28 = 28. These nonstandard math sentences may facilitate learning of mathematical equivalence. If this is the case, then the findings of Experiment 1 could be attributed to the beneficial effect of nonstandard math sentences, rather than to the negative effect of typical arithmetic problems. To rule out this alternative explanation, the control condition in Experiment 2 did not include nonstandard math sentences. Results were consistent with Experiment 1; however, the possibility remains that some other factor varied systematically between the arithmetic and control conditions across experiments.

For example, the stimuli in the arithmetic and control conditions differed in terms of novelty. Children rarely see the equal sign presented by itself or in the context of nonstandard math sentences (Seo & Ginsburg, 2003); therefore, the stimuli presented in the control conditions were less familiar to children than were the stimuli presented in the arithmetic conditions. This novelty alone may have improved children's performance by increasing children's attention to the equal sign either during the lessons or during the equation-solving test, or both. Increasing children's attention to the equal sign has been shown to affect children's strategy choices on math equivalence problems (Alibali, McNeil, & Perrott, 1998). However, increased attention to the equal sign cannot, by itself, account for better performance in the control conditions because prior work has shown no evidence that it affects use of *correct* strategies (Alibali et al., 1998). Another factor that varied between the two conditions was the amount of exposure to numbers. There were more numbers in the arithmetic conditions than in the control conditions. It is possible that the preponderance of numbers increased children's attention to the numbers at the expense of other important features of the lessons. This increased attention to the numbers also may have hurt performance by carrying over to the equation-solving test. If children attended only to the numbers and ignored the location of the equal sign, then they would not have been able to solve the math equivalence problems correctly. Obviously, there is much work to be done before we fully understand the processes involved in the observed effects. Nonetheless, the current study is valuable because it provides the first evidence that children benefit less from a lesson on mathematical equivalence in the presence of typical arithmetic problems than in the presence of either nonstandard math sentences or the equal sign alone. This finding has potential implications for mathematics classrooms.

Educational Implications

The current study adds to the debate about the most effective ways to teach children new information that diverges from their well-established knowledge. Nathan et al. (2002) argue that the most effective instructions are ones that "bridge" from children's well-established knowledge to new material (see also Bryce & MacMillan, 2005; Clement et al., 1989; Koedinger & Nathan, 2004), and some researchers suggest that it is especially beneficial to activate children's incorrect knowledge during instruction because it provides an opportunity for cognitive conflict (Limon, 2001; Piaget, 1980; VanLehn, 1996). However, the present results suggest that it may not always be a good idea to activate children's well-established knowledge. In some cases, the activation of well-established knowledge may hinder children's learning of new information. Teachers may want to avoid activating well-established knowledge in these cases. Specifically, teachers may want to avoid activating children's well-established knowledge of typical arithmetic problems when teaching higher order topics such as mathematical equivalence.

If the activation of children's knowledge of arithmetic contributes to children's difficulties learning mathematical equivalence, then it may be the case that children who are most proficient with typical arithmetic problems are slowest at developing an understanding of algebraic concepts. Consistent with this view, children with the strongest representations of the typical arithmetic patterns have the most difficulties learning from a one-on-one lesson on mathematical equivalence (McNeil & Alibali, 2005b). This hypothesis conflicts with intuition and traditional mathematics practices. Intuition suggests that children who are best at one math topic should be best at another math topic. And indeed, most schools assign, or track, children to algebra based on their performance in arithmetic. This policy is based on theories that suggest that children need to be proficient with "basic" arithmetic facts to be successful at algebra (e.g., Haverty, 1999; Haverty, Koedinger, Klahr, & Alibali, 2000; Kaye, 1986; McKeown, 2002). However, this policy might not be the best approach if children's overly narrow knowledge of arithmetic hinders learning of algebra. Future research in our lab will examine how proficiency with typical arithmetic problems affects understanding of algebra.

Do results of the present experiments imply that teachers should stop teaching arithmetic? Certainly not, but they do suggest that teachers may want to consider alternative approaches to teaching arithmetic. For example, instead of reintroducing arithmetic procedures every year and restricting algebra to a course in eighth or ninth grade, educators could integrate more algebraic ways of thinking into elementary school mathematics, starting in first grade as recommended by many math education researchers (e.g., Blanton & Kaput, 2003; Carpenter & Levi, 1999; National Council of Teachers of Mathematics, 2000). One strategy for doing this is to focus instruction on mathematical equivalence (Carpenter et al., 2003). For example, instead of simply practicing typical arithmetic "facts" such as "3 + 4 = 7," children can learn "7 = 3 + 4," "3 + 4 = 5 + 2," and "7 = 7." Approaches such as this may prevent the entrenchment of overly narrow knowledge of arithmetic in the first place and facilitate the notoriously difficult transition from arithmetic to algebra. We are currently testing the benefits of practice with nonstandard arithmetic problems.

Critics may argue that the entrenchment of overly narrow knowledge of arithmetic is not an important educational problem because performance on math equivalence problems tends to improve after age 9 (McNeil, 2007). In line with this view, the fifth-grade children in Experiment 2 performed reasonably well on math equivalence problems after only receiving a brief intervention. However, it may be possible for older children to improve their performance on math equivalence problems without developing a sophisticated understanding of mathematical equivalence. Indeed, children can learn to perform well on problems despite substantial limitations in their conceptual knowledge (Sophian, 1997). We know that representations of the overly narrow operational patterns continue to coexist along side correct understandings of mathematical equivalence in middle school and beyond (McNeil & Alibali, 2005a, 2005b; Experiment 2). The activation of these overly narrow representations may contribute to many of the difficulties that older students have when working with symbolic expressions and equations in algebra (Carpenter et al., 2003; Kieran, 1981; Knuth et al., 2006). For example, failure to interpret the equal sign relationally may contribute to middle school students' difficulties solving algebraic equations such as $4m + 10 = 70$ (Knuth et al., 2006). Similarly, commitment to the "operations = answer" format may contribute to high school students' difficulty understanding algebraic expressions that do not contain an equal sign such as $3a$, $a + 3$, and $3a + 5a$ (Kieran, 1981). For these reasons among others, it is important to study the best ways to help children overcome their overly narrow views of arithmetic.

Methodological Advancement

One of the primary contributions of the present study is a new methodology for studying factors that affect children's learning. Since the Education Sciences Reform Act of 2002, the establishment of the Institute of Education Sciences, and the publication of *Scientific Research in Education* (National Research Council, 2002), researchers have increased their efforts to apply experimental methods to the study of educational problems. However, it is difficult to perform rigorously controlled experiments in schools. At the same time, educators are highly skeptical of rigorously controlled experiments that occur outside of schools (in laboratories) because they are often small in scale and do not involve the complexities encountered in real-life educational settings. The design used in Experiment 1 provides a way to implement the rigorous control of a laboratory experiment within the context of a classroom.

In Experiment 1, lesson booklets were used to randomize children within classrooms to conditions, and teachers were given a script that applied to both conditions. This method allowed researchers to have a high degree of control over what children saw and heard during the study (similar to the laboratory setting). At the same time, teachers administered the lessons and collected all measures in a typical classroom setting. Of course, critics of classroom-based *experiments* may argue that the booklet format provides a shallow imitation of the rich lessons that teachers should be implementing in the classroom, and critics of *classroom-based* research may argue that the experiment was not as systematic and controlled as it would have been if it were performed one-on-one in the laboratory. Given the large gulf between these two camps, the booklet methodology seems to offer a good compromise. Researchers who are interested in children's learning may want to consider including within-classroom randomization in their methodological toolkit.

Concluding Remarks

The present findings contribute to our understanding of the processes involved in children's failures to learn in seemingly straightforward situations. Such failures historically have been attributed to general conceptual limitations in childhood. However, the current study adds to a growing literature that emphasizes the role of domain knowledge and basic learning processes in children's failures. Results suggest that children's learning failures may be caused, at least in part, by the activation of well-established knowledge. Because many studies focus on how knowledge helps learning (e.g., Rittle-Johnson, Siegler, & Alibali, 2001), situations in which knowledge hinders learning provide a unique window into how the mind works. Future research efforts should be devoted to developing a general framework for understanding when knowledge helps versus hinders learning. By studying how existing knowledge plays a role in the acquisition of new information, researchers can inform both our understanding of how humans acquire knowledge and our efforts to modify the external environment to enhance children's learning and development.

References

Alibali, M. W. (1999). How children change their minds: Strategy change can be gradual or abrupt. *Developmental Psychology, 35*, 127–145.

Alibali, M. W., McNeil, N. M., & Perrott, M. A. (1998). What makes children change their minds? Changes in encoding lead to changes in strategy selection. In M. A. Gernsbacher & S. Derry (Eds.), *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society* (pp. 36–41). Mahwah, NJ: Erlbaum.

Baroody, A. J., & Ginsburg, H. P. (1983). The effects of instruction on children's understanding of the "equals" sign. *Elementary School Journal, 84*, 199–212.

Barsalou, L. W. (1982). Context-independent and context-dependent information in concepts. *Memory and Cognition, 10*, 82–93.

Beaton, A. E., Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., Kelly, D. L., & Smith, T. A. (1996). *Mathematics achievement in the middle school years: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: Boston College.

Behr, M., Erlwanger, S., & Nichols, E. (1980). How children view the equal sign. *Mathematics Teaching, 92*, 13–15.

Blanton, M. L., & Kaput, J. J. (2003). Developing elementary teachers' "algebra eyes and ears." *Teaching Children Mathematics, 10*, 70–77.

Bruner, J. S. (1957). On perceptual readiness. *Psychological Review, 2*, 123–152.

Bryce, T., & MacMillan, K. (2005). Encouraging conceptual change: The use of bridging analogies in the teaching of action-reaction forces and the "at rest" condition in physics. *International Journal of Science Education, 27*, 737–763.

Carpenter, T. P., Franke, M. L., & Levi, L. (2003). *Thinking mathematically: Integrating arithmetic and algebra in elementary school*. Portsmouth, NH: Heinemann.

Carpenter, T. P., & Levi, L. (1999, April). *Developing conceptions of algebraic reasoning in the primary grades*. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.

Clement, J., Brown, D., & Zeitsman, A. (1989). Not all preconceptions are misconceptions: Finding anchoring conceptions for grounding instruction on students' intuition. *International Journal of Science Education, 11*, 554–565.

Cobb, P. (1987). An investigation of young children's academic arithmetic contexts. *Educational Studies in Mathematics, 18*, 109–124.

Collis, K. F. (1974). Cognitive development and mathematics learning. Paper presented at the Psychology of Mathematics Education Workshop, Centre for Science Education, Chelsea College, London.

Haverty, L. A. (1999). *The importance of basic number knowledge to advanced mathematical problem solving*. Unpublished doctoral dissertation, Carnegie Mellon University, Pittsburgh.

Haverty, L. A., Koedinger, K. R., Klahr, D., & Alibali, M. W. (2000). Solving inductive reasoning problems in mathematics: Not so trivial pursuit. *Cognitive Science, 24*, 249–298.

Jensen, R. J. (1990). The spring is wound too tight on our spiral curriculum. *Arithmetic Teacher, 38*, 4–5.

Kaye, D. B. (1986). The development of mathematical cognition. *Cognitive Development, 1*, 157–170.

Kieran, C. (1981). Concepts associated with the equality symbol. *Educational Studies in Mathematics, 12*, 317–326.

Kieran, C. (1989). The early learning of algebra: A structural perspective. In S. Wagner & C. Kieran (Eds.), *Research issues in the learning and teaching of algebra* (pp. 33–56). Reston, VA: NCTM.

Knuth, E. J., Stephens, A. C., McNeil, N. M., & Alibali, M. W. (2006). Does understanding the equal sign matter? Evidence from solving equations. *Journal for Research in Mathematics Education, 37*, 297–312.

Koedinger, K. R., & Nathan, M. J. (2004). The real story behind story problems: Effects of representation on quantitative reasoning. *Journal of the Learning Sciences, 13*, 129–164.

Langer, E. J. (2000). Mindful learning. *Current Directions in Psychological Science*, *9*, 220–223.

Li, X., Ding, M., Capraro, M. M., & Capraro, R. M. (2008). Sources of differences in children's understandings of mathematical equality: Comparative analysis of teacher guides and student texts in China and the United States. *Cognition and Instruction*, *26*, 195–217.

Limon, M. (2001). On the cognitive conflict as an instructional strategy for conceptual change: A critical appraisal. *Learning and Instruction*, *11*, 357–380.

McGilly, K., & Siegler, R. S. (1990). The influence of encoding and strategic knowledge on children's choices among serial recall strategies. *Developmental Psychology*, *26*, 931–941.

McKeown, M. (2002). *Does two plus two still equal four: What should our children know about math?* Talk presented at the American Enterprise Institute for Public Policy Research (AEI). Retrieved June 2005 from http://www.aei.org/events/eventID.193/event_detail.asp

McNeil, N. M. (2005). *When the basics backfire: The change-resistance account of equation-learning difficulties*. Unpublished doctoral dissertation, University of Wisconsin, Madison.

McNeil, N. M. (2007). U-shaped development in math: Seven year olds outperform nine year olds on equivalence problems. *Developmental Psychology*, *43*, 687–695.

McNeil, N. M., & Alibali, M. W. (2000). Learning mathematics from procedural instruction: Externally imposed goals influence what is learned. *Journal of Educational Psychology*, *92*, 734–744.

McNeil, N. M., & Alibali, M. W. (2004). You'll see what you mean: Students encode equations based on their knowledge of arithmetic. *Cognitive Science*, *28*, 451–466.

McNeil, N. M., & Alibali, M. W. (2005a). Knowledge change as a function of mathematics experience: All contexts are not created equal. *Journal of Cognition and Development*, *6*, 285–306.

McNeil, N. M., & Alibali, M. W. (2005b). Why won't you change your mind? Knowledge of operational patterns hinders learning and performance on equations. *Child Development*, *76*, 1–17.

McNeil, N. M., Grandau, L., Knuth, E. J., Alibali, M. W., Stephens, A. C., Hattikudur, S., et al. (2006). Middle-school students' understanding of the equal sign: The books they read can't help. *Cognition and Instruction*, *24*, 367–385.

Nathan, M. J., Stephens, A. C., Masarik, D. K., Alibali, M. W., & Koedinger, K. R. (2002). Representational fluency in middle school: A classroom based study. In P. Mewborn, P. Sztajn, D. White, H. Wiegel, R. Bryant, & K. Nooney (Eds.), *Proceedings of the XXIV Annual Conference of the North American Chapter of the International Group for the Psychology of Mathematics Education* (Vol. I, pp. 463–472). Columbus, OH: ERIC.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

National Research Council. (2002). *Scientific research in education*. Washington, DC: National Academy Press.

Perry, M. (1991). Learning and transfer: Instructional conditions and conceptual change. *Cognitive Development*, *6*, 449–468.

Perry, M., Church, R. B., & Goldin-Meadow, S. (1988). Transitional knowledge in the acquisition of concepts. *Cognitive Development*, *3*, 359–400.

Piaget, J. (1980). *Experiments in contradiction*. Chicago: University of Chicago Press.

Piaget, J., & Szeminska, A. (1995). The child's conception of number (C. Gattegno & F. M. Hadgson, Trans.). In H. E. Gruber & J. J. Voneche (Eds.), *The essential Piaget* (pp. 298–341). Northvale, NJ: Jason Aronson. (Original work published 1941)

Rittle-Johnson, B., & Alibali, M. W. (1999). Conceptual and procedural knowledge of mathematics: Does one lead to the other? *Journal of Educational Psychology*, *91*, 175–189.

Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology*, *93*, 346–362.

Schauble, L. (1990). Belief revision in children: The role of prior knowledge and strategies for generating evidence. *Journal of Experimental Child Psychology*, *49*, 31–57.

Seo, K.-H., & Ginsburg, H. P. (2003). "You've got to carefully read the math sentence.": Classroom context and children's interpretations of the equals sign. In A. J. Baroody & A. Dowker (Eds.), *The development of arithmetic concepts and skills* (pp. 161–187). Mahwah, NJ: Erlbaum.

Sherman, J. L. (2007). *Children's misunderstanding of the equal sign and their adherence to addition schemas: Detecting the effect of prior arithmetic experience and instruction*. Unpublished doctoral dissertation, University of Alberta, Edmonton, Alberta, Canada.

Siegler, R. S. (2000). The rebirth of children's learning. *Child Development*, *71*, 26–35.

Sophian, C. (1997). Beyond competence: The significance of performance for conceptual development. *Cognitive Development*, *12*, 281–303.

VanLehn, K. (1996). Cognitive skill acquisition. *Annual Review of Psychology*, *47*, 513–539.

Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development*, *11*, 37–63.

Child Development, September/October 2008, Volume 79, Number 5, Pages 1538–1554

Perception of Dynamic and Static Audiovisual Sequences in 3- and 4-Month-Old Infants

David J. Lewkowicz
Florida Atlantic University

This study investigated perception of audiovisual sequences in 3- and 4-month-old infants. Infants were habituated to sequences consisting of moving/sounding or looming/sounding objects and then tested for their ability to detect changes in the order of the objects, sounds, or both. Results showed that 3-month-olds perceived the order of 3-element sequences and that 4-month-olds perceived the order of 3- and 4-element sequences. In addition, results indicated that perception of moving sequences was more robust than of looming ones and that detection of order differences was not based on primacy cues. The current results show that perception of multisensory sequences and their structure emerges early in development, and a preliminary model of the development of sequence perception is offered.

The ability to perceive, learn, and produce sequences provides the foundation for many behavioral functions (Keele, Ivry, Mayr, Hazeltine, & Heuer, 2003; Lashley, 1951; Martin, 1972; Zacks & Tversky, 2001). For example, our ability to tell that the utterance "the boy kissed the girl" differs from the utterance "the girl kissed the boy" depends on the fact that we can tell that the same words are ordered differently in the two phrases. Likewise, our ability to hear two differently ordered sequences of identical musical notes as different melodies depends on our ability to detect the ordering of the notes. Finally, our ability to execute a series of hand and arm movements when driving a car depends on the fact that we are able to assemble the required component actions into the appropriately ordered sequence. These examples make it clear that our ability to detect the particular ordering of a series of sequence elements is key to discovering the "hidden" meanings inherent in different types of events and that our ability to concatenate a series of motor acts into a correctly ordered action sequence is key to producing adaptive behaviors.

Given the fundamental importance of the ability to perceive, learn, and produce sequences, there is little doubt that understanding its development can provide critical insights into the development of basic perceptual, cognitive, and social skills (Greenfield, 1991; Nelson, 1986). Piaget (1952) recognized this fact and conducted careful observations to determine whether and when infants might be able to produce sequences of behavioral actions to achieve specific goals. He found that infants as young as 8 months of age could string together a series of discrete motor acts into adaptive action sequences. More recent empirical investigations have supported Piaget's observations by showing that the ability to produce action sequences emerges early in life. Specifically, these investigations have shown that (a) 4-month-old infants can engage in turn taking when interacting with adults (Jaffe, Beebe, Feldstein, Crown, & Jasnow, 2001); (b) 3-month-old infants can learn to move a series of three mobiles in a particular sequence and can remember the order in which they moved them 24 hr later (Merriman, Rovee-Collier, & Wilk, 1997); (c) 6-month-old infants can learn to move a series of as many as five mobiles in a particular sequence (Gulya, Sweeney, & Rovee-Collier, 1999); (d) 9-month-old infants can learn, remember, and reproduce multiact manual sequences in the correct sequential order as long as the acts are related in obvious ways (Bauer, Wiebe, Waters, & Bangston, 2001; Carver & Bauer, 1999); and (e) 22-month-old children can learn and remember multiact sequences even when the acts are arbitrarily related (Bauer, Hertsgaard, Dropik, & Daly, 1998).

A number of other studies have investigated whether infants can perceive and learn sequences. Like the studies on sequence production, these

---

I thank Marcia Dabbene and Jennifer Hughes for their assistance in data collection. Some of the data for this article were collected while the author was at the New York State Institute for Basic Research in Developmental Disabilities in Staten Island, NY. This work was supported in part by the New York State Office of Mental Retardation and Developmental Disabilities and by the National Institute of Child Health and Human Development Grant R01 HD35849.

Correspondence concerning this article should be addressed to David J. Lewkowicz, Department of Psychology, Florida Atlantic University, 777 Glades Road, Boca Raton, FL 33431. Electronic mail may be sent to lewkowic@fau.edu.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7905-0021

studies have found that infants can, indeed, perceive and learn sequences. Specifically, they have shown that (a) 5-month-old infants can learn and then recall the sequential arrangement of a series of four static images (Smith, Loboschefski, Davidson, & Dixon, 1997), (b) 8-month-old infants can learn statistical (i.e., transitional probability) relations among adjacent sounds (Saffran, Aslin, & Newport, 1996; Saffran, Johnson, Aslin, & Newport, 1999), (c) 2- to 8-month-old infants can learn statistical relations among adjacent visual stimuli (Kirkham, Slemmer, & Johnson, 2002), (d) 9-month-old infants prefer speech strings and monosyllabic nonwords that follow legal phonotactic adjacency rules (Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993; Jusczyk, Luce, & Charles-Luce, 1994), (e) 7- to 12-month-old infants can learn simple grammatical rules for sequences of sounds (Gómez & Gerken, 1999; Marcus, Vijayan, Rao, & Vishton, 1999), and (f) 15- to 18-month-old infants can learn distant sequential relations specifying a series of sounds (Gómez, 2002; Gómez & Maye, 2005).

Interestingly, when the findings on infant ability to perceive and learn sequences are considered together, they clearly demonstrate that language is not necessary for the development of sequence perception and learning. Indeed, this conclusion is supported by findings from studies of nonhuman primates showing that monkeys also not only can learn sequences but that they can generalize across sequences composed of different elements and, thus, can learn ordinal position information as well as syntax-like rules (Chen, Swartz, & Terrace, 1997; Hauser, Weiss, & Marcus, 2002; Orlov, Yakovlev, Hochstein, & Zohary, 2000; Terrace, 1993; Terrace, Son, & Brannon, 2003).

Overall, the extant empirical findings indicate that the ability to produce organized action sequences and perceive sequential relations emerges very early in life, that it improves rapidly in early development, and that infants' ability to perceive sequences is quite general because they can perceive and learn sequences specified in the auditory as well as the visual modalities. Given these facts, it is reasonable to suppose that the early emerging ability to perceive sequences contributes in important ways to the development of higher level perceptual and cognitive skills. The most likely way in which it does so is by providing the necessary early perceptual and developmental foundation for the subsequent discovery of the hidden sequential structure underlying most events. For example, in language development, infants probably initially discover the ordered and then syntactic structure underlying sequences of sounds, and once they do so, they can then proceed to discover the hidden meanings that are inherent in such sequences of sounds. If such a progressive developmental scenario is general in nature, then a similar pattern might be observed in the development of event perception. That is, infants probably initially discover the ordered and then causal relations among sequence elements, and once they do so, they then can proceed to discover the hidden meanings inherent in particular events.

It is interesting to note that most of the events in an infant's daily life are specified by multisensory attributes that are typically redundant. For example, when infants see and hear a person talking, or when they see and hear a ball bouncing, they have access to synchronous auditory and visual sensory attributes that correspond in terms of their duration, tempo, and rhythmic patterning. It is also interesting to note that multisensory redundancy increases perceptual salience and facilitates perception, learning, and discrimination in human adults (Hershenson, 1962; McGurk & MacDonald, 1976; Summerfield, 1979; Welch & Warren, 1980), animal adults (Partan & Marler, 1999; Rowe, 1999; Stein & Meredith, 1993), and human infants (Bahrick & Lickliter, 2000; Lewkowicz & Kraebel, 2004; Neil, Chee-Ruiter, Scheier, Lewkowicz, & Shimojo, 2006; Scheier, Lewkowicz, & Shimojo, 2003). As a result, it is important to investigate infant response to multisensory event sequences if we are to gain deeper insights into the development of sequence perception and learning.

Only one study to date has investigated infant response to multisensory sequences specified by auditory and visual attributes (Lewkowicz, 2004) and, as might be expected, found that multisensory redundancy did, in fact, enhance young infants' response. Specifically, in this study, infants were habituated to a sequence of three moving objects and their concurrent impact sounds and then were tested in separate test trials to determine whether they could detect changes in the serial order of the objects, the impact sounds alone, or the objects and their concurrent impact sounds. Results indicated that 4-month-old infants detected serial order changes only when they were specified concurrently by audible and visible attributes but that 8-month-old infants detected the changes regardless of whether they were specified concurrently by the audible and visible attributes or by just one of them.

Although the Lewkowicz (2004) study provided evidence that infants can learn and discriminate dynamic audiovisual sequences, it also raised some key questions. First, a general sensitivity to temporal and dynamic information—the essential underlying basis for learning dynamic audiovisual sequences—emerges earlier than 4 months of age (Kellman &

Arterberry, 1998; Lewkowicz, 1989, 2000), thus raising the question of whether infants younger than 4 months of age might be able to learn audiovisual sequences. Second, the sequences presented in the Lewkowicz (2004) study differed in terms of their initial objects/sounds, raising the question of whether this difference made discrimination possible. Third, because sensitivity to visual motion emerges very early in infancy (Bertenthal, Proffitt, Kramer, & Spetner, 1987; Dannemiller & Freedland, 1991; Kaufmann-Hayoz, Kaufmann, & Stucki, 1986; Kaufmann, Stucki, & Kaufmann-Hayoz, 1985; Shea & Aslin, 1990) and because motion contributes to object perception in infancy (Kellman & Spelke, 1983; Kellman, Spelke, & Short, 1986), this raises the question of whether motion might play a role in infant perception of dynamic audiovisual sequences. Finally, studies have shown that infants are sensitive to visual sequence complexity (Canfield & Haith, 1991; Gulya, Rovee-Collier, Galluccio, & Wilk, 1998; Gulya et al., 1999), thus raising the question of whether infants also are sensitive to audiovisual sequence complexity?

The purpose of the current study was to answer the four questions outlined above in separate experiments. Thus, Experiment 1 investigated whether 3-month-old infants can perceive and discriminate dynamic audiovisual sequences by presenting the identical stimuli and using the same testing procedures as in the Lewkowicz (2004) study. Experiment 2 investigated whether primacy effects might contribute to discrimination of audiovisual dynamic sequences by presenting contrasting sequences whose initial sequence elements were held constant. Experiment 3 investigated whether motion may contribute to perception and discrimination of dynamic audiovisual sequences by presenting sequences consisting of looming/sounding objects. Finally, Experiment 4 examined the effects of sequence length/complexity by testing infants' ability to perceive and discriminate four-element sequences.

Experiment 1

Studies have shown that infants younger than 4 months of age can learn sequences. For example, Canfield and Haith (1991) presented sequences consisting either of symmetrically or of asymmetrically alternating visual stimuli and found that 2-month-old infants learned to anticipate symmetrical sequences and that 3-month-old infants learned to anticipate asymmetrical sequences as well. Gulya et al. (1998) used the mobile conjugate reinforcement procedure to train 3-month-old infants to move a series of different mobiles and found that infants could learn to move a series of three mobiles and then remember the order in which they moved them a day later. Finally, Kirkham et al. (2002) found that infants as young as 2 months of age could quickly learn the statistical relations specifying a series of visual shapes. When these findings are considered together with the fact that sensitivity to temporal information and to visual motion emerges prior to 4 months of age (Lewkowicz, 1989, 2000), they suggest that infants younger than 4 months of age should be able to perceive, learn, and discriminate dynamic audiovisual sequences. The purpose of this experiment was to test this prediction by investigating 3-month-old infants' response to the same dynamic audiovisual sequences presented by Lewkowicz (2004) to 4- and 8-month-old infants.

Method

Participants. Twenty-four 3-month-old infants (M age = 14.7 weeks, SD = 0.7 weeks; 12 boys and 12 girls) contributed data in this experiment. Three additional infants were tested but were not used because they were inattentive. The infants in this experiment, as well as in all the subsequent ones, were full term at the time of birth, had birth weights above 2,500 g, had 1- and 5-min Apgar scores of 7 or higher, had no prenatal or perinatal complications, and were healthy at the time of testing. The ethnic proportions of the participants in all the experiments reported here were approximately 5% Hispanic, 77% non-Hispanic, and 18% unreported. The corresponding racial proportions were 4% Black or African American, 0% Asian, 3% more than one race, 73% White or Caucasian, and 20% unreported.

Apparatus and stimuli. All stimulus events were multimedia movies. One movie showed an attention-getter, which was an expanding and contracting green disk. A second movie, used to test for possible fatigue effects, showed a segment of a Winnie the Pooh cartoon. It was presented at a sound pressure level (SPL) of 70-74 dB against an ambient SPL of 50 dB. The remaining four movies showed different sequential orderings of three different moving objects and their distinct impact sounds. Each of these movies showed the three objects seen in Figure 1a (Object A—circle, Object B—triangle, and Object C—square) emerging one after the other from a spout at the top of the stimulus presentation monitor. The objects moved down, and when they reached the bottom of their downward path, they contacted a black ramp (part of which was not visible in this experiment), made an impact sound, turned to the right, and moved off to

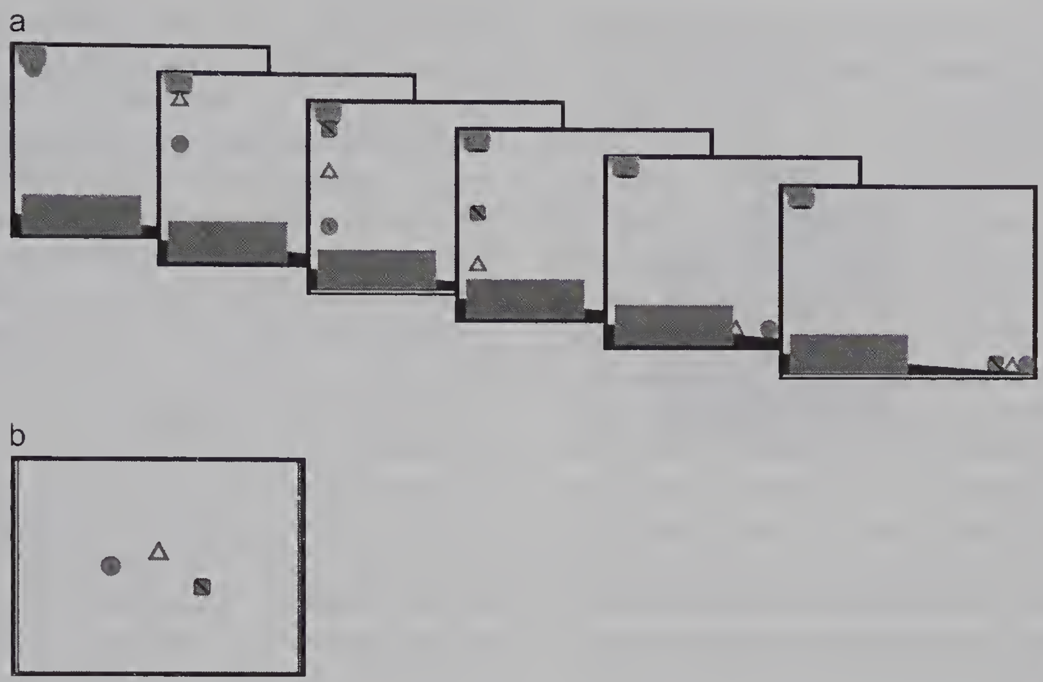


Figure 1. (a) The three objects and the schematic representation of their movement over time in Experiment 1. (b) The spatial arrangement of the three objects in Experiment 3.

the side (the objects did not pause when they made their impact sound). The impact sounds were ".wav" files that are part of the Microsoft Windows 98 operating system and corresponded to the three objects as follows: Object A—JungleSingDrumMono.wav, Object B—UtopiaDink.wav; and Sound C—SportsPPongMono.wav. The movies consisting of the moving/sounding objects represented the four possible combinations of object and/or sound ordering. They were as follows: auditory-ABC/visual-ABC, auditory-CAB/visual-ABC, auditory-ABC/visual-CAB, and auditory-CAB/visual-CAB. All movies were presented on a 17-in. computer monitor at a distance of 50 cm from the infant. At this distance, the objects subtended a mean of 2.86° of visual angle (the angular size is reported as an average because the objects varied very slightly in their size). The audio part of the movie was presented through two speakers placed on each side of the monitor. The mean SPL of the impact sounds was 80 dB (A scale) against an ambient SPL of 50 dB.

As shown in Figure 1a, each movie began with the appearance of the spout, the ramp, and the gray rectangle placed in front of the ramp. As soon as the spout, the ramp, and the gray rectangle appeared, the three objects began to emerge from the spout at half-second intervals and moved down at the same and constant speed. As each object moved down, it disappeared behind the gray rectangle on its way down, made an impact sound while still not visible, turned to the right, moved down the ramp, and finally emerged from behind the gray rectangle and moved off to the right where it came to rest. It took 1.55 s for each object to reach the ramp and make its impact sound. It took another 3.6 s for each object to reach the end of its trajectory at the right of the screen. Once the last object came to rest, all three objects remained visible for 0.5 s, disappeared for 0.75 s, and then the sequence started again and continued to be presented repeatedly either until the infant looked away or until the maximum trial duration was reached (see below).

Procedure. An experimenter observed the infant from a separate control room and, thus, could neither see nor hear the stimuli being presented. Because we used the infant-controlled habituation/test procedure, the infant's looking behavior controlled movie presentation. Whenever the infant looked at the stimulus presentation monitor, the movie began to play and continued to do so either until the infant looked away from the monitor for more than 1 s or until he or she accumulated a total of 48 s looking time. Once either of these two conditions was met, the attention-getter reappeared and remained there until the infant looked back at the monitor. As soon as the infant refixated the attention-getter, the next trial was initiated.

The experiment began with a single pretest trial during which the Winnie the Pooh cartoon was played. The purpose of this trial was to measure the infant's initial level of attention. Once this trial ended, the habituation phase began and continued until the infant reached a predefined habituation criterion. This criterion required that the total duration of looking during the last three habituation trials decline to 50% of the total duration of looking during the first three habituation trials. Once an infant reached the habituation criterion, the habituation phase ended and the next trial initiated the test phase.

Infants were divided into two habituation groups. One group was habituated to the auditory-ABC/visual-ABC movie and the other to the auditory-CAB/visual-CAB movie. Once infants reached the habituation criterion, they were given four test trials. The first of these was a familiar (F) test trial during which the movie that was presented during the habituation phase was presented again. Looking during this trial served as the baseline measure of responsiveness against which responsiveness to the order changes presented in the remaining test trials was assessed. The three remaining test trials were designed to test infants' response to a change in the order of the auditory, the visual, or the audiovisual attributes of order and were designated the A, V, and AV test trials, respectively. The A test trial involved a change in the order of the sounds but not in the order of the objects, the V test trial involved a change in the order of the objects but not in the order of the sounds, and the AV test trial involved a change in the order of sounds and objects. The F test trial was always presented first, whereas the order of the A, V, and AV test trials was counterbalanced according to a Latin square design. Once the last test trial was completed, each infant was given a posttest trial during which the Winnie the Pooh cartoon was presented. This trial served as a check on fatigue effects by measuring the terminal level of attention. Once this trial was completed, the test session ended.

Results and Discussion

An initial analysis was performed to determine whether any infants exhibited spontaneous regression to the mean in the F test trial. Spontaneous regression to the mean was deemed to have occurred for a given infant if his or her looking duration score in the F test trial exceeded the overall mean duration of looking in the F trial by more than 2 *SD*. One infant exhibited spontaneous regression to the mean, and this infant's data were eliminated from any further analyses. Figure 2 shows the results from the habituation and test trials for the remaining 23 infants. The left panel of Figure 2 shows the data from the first three and last three habituation trials and shows that attention declined during the habituation phase (only the first and last three trials are shown because the infant-controlled procedure permits a variable number of habituation trials across infants). To determine whether the decline was statistically reliable and whether it might have differed for the two habituation groups, the data were analyzed by way of a mixed 2 × 6 (Habituation Group × Habituation Trials) repeated measures analysis of variance (ANOVA), with habituation group as the between-subjects factor and habituation trials as the within-subjects factor. Results of this analysis indicated that the habituation trials effect was statistically significant, $F(5, 105) = 20.4$, $p < .001$, and that no other effects were significant. This result indicates that infants exhibited a statistically reliable decrement in responsiveness during the habituation phase and that the nature of this decrement was not affected by the particular sequence presented during habituation.

The right panel of Figure 2 shows the data from the test trials. These data were first analyzed by way of a mixed 2 × 4 (Habituation Group × Test Trial Type) mixed ANOVA, with habituation group as the between-subjects factor and test trial type as a within-subjects factor. Results of this analysis indicated that neither the overall trials effect was significant, $F(3, 63) = 2.28$, $p < .10$, nor any of the other effects. Despite the fact that the overall trials effect was not statistically significant, planned comparisons were theoretically and empirically (Lewkowicz, 2004) justified. Thus, to determine whether infants exhibited statistically reliable response recovery in any of the test trials, separate planned comparison analyses were performed contrasting the duration of looking in each order-change trial, respectively, with the duration of looking in the F test trial. These comparisons indicated that the only test trial in which response recovery reached statistical significance was the AV test trial, $F(1, 21) = 5.26$, $p < .05$. Nonetheless, it is interesting that the average magnitude of response in the A and V test trials was higher than in the F test trial. This raises the possibility that response in these two trials might have reached conventional levels of statistical significance if sample size were increased. It should be noted, however, that the purpose of the current experiment was to faithfully repeat Lewkowicz's (2004) Experiment 2. As a result, it was deemed important that sample size be kept the same as in the original study.

The finding that 3-month-old infants detected the order change in the AV test trial replicates the findings from Lewkowicz's (2004) Experiment 2 and indicates that infants as young as 3 months of age can perceive and discriminate dynamic audiovisual sequences. Second, the fact that 3-month-old infants, like the 4-month-old infants in the Lewkowicz (2004) study, only responded to the audiovisual order change is consistent with the fact that multisensory redundancy is particularly important for infant learning and discrimination in the early months of life (Bahrick, Lickliter, & Flom, 2004). Third, the fact that 3-month-old

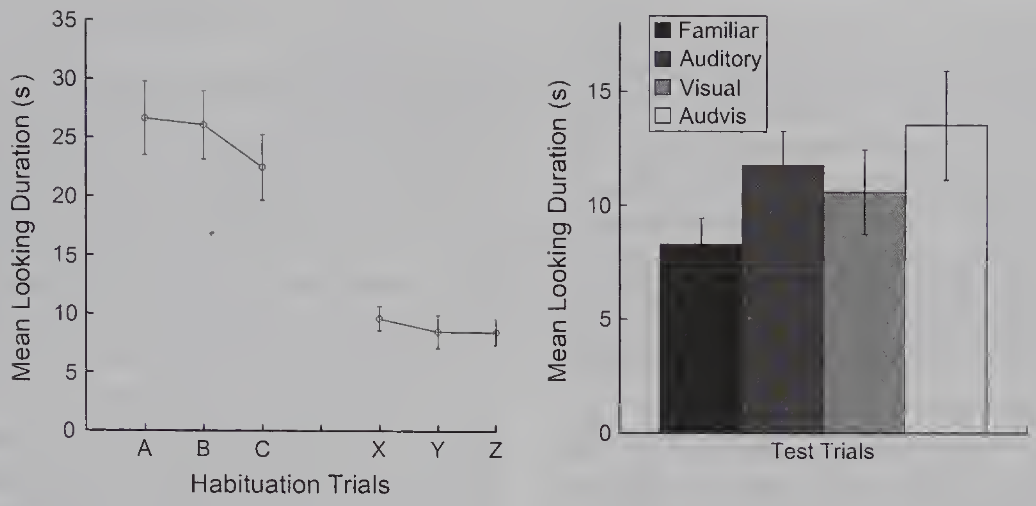


Figure 2. Results from Experiment 1.
Note. The left panel shows the duration of looking during the habituation phase. Because the number of trials to reach the habituation criterion varied across infants, the left panel only depicts the average duration of looking in the first (A, B, and C) and the last three (X, Y, and Z) habituation trials. The right panel shows the duration of looking in the test trials. Error bars indicate the standard error of the mean.

infants perceived and discriminated audiovisual sequences is consistent with previous reports of successful sequence perception in infants of this age (Canfield & Haith, 1991; Gulya et al., 1998) as well as with the fact that temporal- and motion-processing abilities are already present by this age (Lewkowicz, 1989, 2000).

Experiment 2

The contrasting sequences that were presented in Experiment 1 and in Lewkowicz's (2004) study differed in terms of their initial objects and sounds, making it possible that successful discrimination of the contrasting sequences was based on initial element differences. In other words, successful discrimination might have been based on the classic primacy effect rather than on the perception of serial order. It should be noted, however, that this is unlikely for two reasons. First, neither the 3-month-old infants in Experiment 1 nor the 4-month-old infants in Lewkowicz's (2004) study exhibited detection of the visual-only order change despite the fact that it involved the initial object. Second, the infants' failure to respond to the visual-only order change means that their successful detection of the audiovisual order change could not have been based on the visual aspects of this change and, by extension, could not have been based on first object differences either.

Despite the fact that the findings suggest that primacy effects did not contribute to successful discrimination, studies have shown that infants can take advantage of initial sequence element differences (Cornell & Bergstrom, 1983; Merriman et al., 1997). As a result, it is imperative to rule out such primacy effects directly through experimental manipulations of sequence structure. The most convincing way to rule them out is to conduct a direct empirical investigation in which learning of dynamic audiovisual sequences is assessed when the initial elements of contrasting sequences are held constant. As a result, Experiment 2 tested 4-month-old infants' ability to perceive and discriminate dynamic audiovisual sequences whose initial elements were the same.

Method

Participants. Twenty-four 4-month-old infants (*M* age = 19.4 weeks, *SD* = 0.7 weeks; 15 boys and 9 girls) contributed data in this experiment. One additional infant was tested but was not used because of fussing.

Apparatus and stimuli. The apparatus and stimuli in this experiment were identical to those used in Experiment 1 except that here the two contrasting sequences were ABC and ACB.

Procedure. The procedure in this experiment was the same as in Experiment 1 except that here half the infants were habituated to an auditory-ABC/visual-ABC movie and the other half were habituated to an auditory-ACB/visual-ACB movie. Following the habituation phase, each infant was given the four test trials shown in Table 1.

Results and Discussion

Initial examination of the data from the F test trial indicated that two infants exhibited spontaneous regression to the mean, and thus, the data from these infants were excluded from all subsequent analyses. Figure 3 shows the results from the habituation and

Table 1

The Sequences Presented to Each Group in the Habituation and Test Trials in Experiments 2 and 3

Habituation group	Test trials			
	Familiar (F)	Auditory (A)	Visual (V)	Auditory – Visual (AV)
ABC	Auditory-ABC/Visual-ABC	Auditory-ACB/Visual-ABC	Auditory-ABC/Visual-ACB	Auditory-ACB/Visual-ACB
ACB	Auditory-ACB/Visual-ACB	Auditory-ABC/Visual-ACB	Auditory-ACB/Visual-ABC	Auditory-ABC/Visual-ABC

Note. The letters A, B, and C refer to each object and its corresponding impact sound (see the Method section for details).

test trials from the remaining 22 infants. The left panel of Figure 3 shows that infants' looking declined during the habituation phase of the experiment. A mixed 2 × 6 (Habituation Group × Habituation Trials) ANOVA, with habituation group as the between-subjects factor and habituation trials as the within-subjects factor, indicated that the habituation trials effect was statistically significant, $F(5, 100) = 38.23$, $p < .001$, and that no other effects were significant. This result indicates that infants exhibited a statistically reliable decrement in responsiveness during the habituation phase and that the nature of this decrement was not affected by the particular sequence presented during the habituation phase.

The right panel of Figure 3 shows the results of the test trials. As can be seen, infants looked longer in each of the order-change test trials relative to the amount of time they looked in the F test trial. A mixed 2 × 4 (Habituation Group × Test Trial Type) ANOVA, with habituation group as the between-subjects factor and test trial type as a within-subjects factor, indicated that responsiveness differed across the test trials, $F(3, 60) = 3.83$, $p < .025$, and that there were no other significant effects. Planned comparison tests comparing response in each order-change test trial, respectively, versus looking in the F test trial indicated that infants exhibited significant response recovery in the A test trial, $F(1, 20) = 9.67$, $p < .01$; the V test trial, $F(1, 20) = 6.84$, $p < .025$; and the AV test trial, $F(1, 20) = 13.5$, $p < .01$.

The results from this experiment indicate that infants successfully discriminated the sequences despite the fact that the initial sequence element was the same. This finding supports Lewkowicz's (2004) conclusion that initial element differences did not account for 4-month-old infants' successful response. Paradoxically, however, the present findings also showed that in contrast to the findings from Experiments 1 and 2 in Lewkowicz's study, here infants detected unisensory as well as bisensory order changes. The absence of a multisensory redundancy effect is actually consistent with other findings on multisensory redundancy effects in that such effects are assumed to depend, in part, on information complexity (Bahrick et al., 2004). The discriminative task in the current experiment was simpler because infants only had to detect a change in the order of two of three sequence elements, whereas in Lewkowicz's study, infants had to detect a change in the order of three elements.

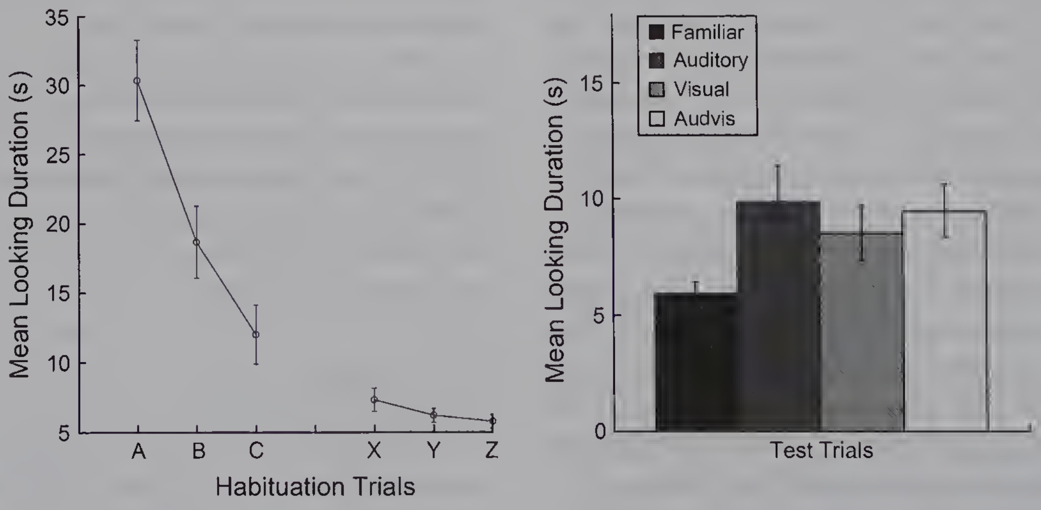


Figure 3. Results from Experiment 2.
Note. The left panel shows the duration of looking during the first three (A, B, and C) and last three (X, Y, and Z) habituation trials, and the right panel shows the duration of looking in the test trials. Error bars indicate the standard error of the mean.

Experiment 3

Several lines of work suggest that motion is likely to contribute in an important way to infant perception and discrimination of dynamic audiovisual sequences. First, findings indicate that infants are sensitive to motion from a very early age (Bertenthal et al., 1987; Dannemiller & Freedland, 1991; Kaufmann et al., 1985; Kaufmann-Hayoz et al., 1986; Shea & Aslin, 1990). Second, evidence indicates that motion plays an important role in object perception in that infants can only perceive the partially visible parts of a moving object as belonging to the same object if the parts move together (Kellman & Spelke, 1983; Kellman et al., 1986). Third, young infants can discriminate between an upright and an upside-down moving point-light display of a walking person on the basis of the correlated motion of the individual lights (Bertenthal, Proffitt, & Cutting, 1984). Fourth, when infants see a moving object, they can anticipate its future position and correctly intercept and catch it (von Hofsten, 1980). Finally, young infants can perceive the temporal relation between a single moving object and its impact sound (Lewkowicz, 1996) and can choose between two moving objects on the basis of their spatiotemporal relationship to an impact sound (Lewkowicz, 1992a).

Despite the possibility that motion is likely to contribute in important ways to infant perception of sequentially structured events, it is currently not known whether it does. This is because, with the exception of the Lewkowicz (2004) study and Experiments 1 and 2 in the current study, all other studies to date on infant sequence perception and learning have presented sounds, looming visual stimuli, or static visual stimuli. Given that young infants are sensitive to motion and that they rely on it for object perception and action, it is interesting to assess the possible contribution of motion to infant perception and learning of dynamic audiovisual sequences. To do so, the present experiment examined 4-month-old infants' response to dynamic sequences whose motion characteristics were reduced (i.e., the objects that loomed and sounded in place).

Method

Participants. Twenty-four 4-month-old infants (*M* age = 19 weeks, *SD* = 0.9 weeks; 15 boys and 9 girls) contributed data in this experiment. An additional 15 infants were tested but were not used because 11 were fussy, 3 were inattentive, and 1 was distracted.

Apparatus and stimuli. The apparatus and stimuli were identical to those used in Experiment 1 except that here all three objects were visible from the beginning to the end of each trial and remained in one place throughout the trial. Figure 1b shows the objects' spatial arrangement. The sequential nature of the event was determined by the objects' looming/sounding behavior. At the start of a trial, the first object gradually increased in size by 20%. When it reached its maximum size, its corresponding auditory stimulus was sounded and the object then returned to its original size. This was followed by the looming of the second and then third object. The looming behavior of each object was continuous in nature, and it took 0.5 s for each object to loom to its maximum size and another 0.5 s to return back to its original size. A 0.5-s interval separated the looming action of each object, and during this interval, all objects remained static and at their original size. Once the last object loomed, all three objects could be seen at their original size for 1 s. This cycle of looming actions was repeated for the duration of the trial.

Procedure. The procedure used in Experiment 1 and the design depicted in Table 1 were used here.

Results and Discussion

Analysis of the F test trial data indicated that two infants exhibited spontaneous regression to the mean in this trial, and as a result, the data from these two infants were eliminated from all subsequent analyses. The habituation and test data from the remaining 22 infants can be seen in Figure 4, and as can be seen in the left panel of Figure 4, looking declined during the habituation phase. A mixed 2 × 6 (Habituation Group × Habituation Trials) ANOVA, with habituation group as the between-subjects factor and habituation trials as the within-subjects factor, indicated that infants exhibited a statistically reliable decrement in responsiveness over trials, $F(5, 100) = 14.12$, $p < .001$, and that the nature of this decrement was not affected by the particular sequence presented during the habituation phase.

The right panel of Figure 4 shows the mean duration of looking in the test trials. A mixed 2 × 4 (Habituation Group × Test Trial Type) ANOVA, with habituation group as the between-subjects factor and test trial type as the within-subjects factor, indicated that responsiveness differed across the test trials, $F(3, 60) = 3.43$, $p < .05$, but that it was not affected by the particular habituation sequence. Planned comparisons indicated that infants exhibited significant response recovery in the A test trial, $F(1, 20) = 6.67$, $p < .025$, but that they did not exhibit response recovery in the other two test trials. As was the case in Experiment 1, here the average magnitude of response in the V and the AV test trials also was higher than in the F test trial, raising the possibility that response in these two trials might have

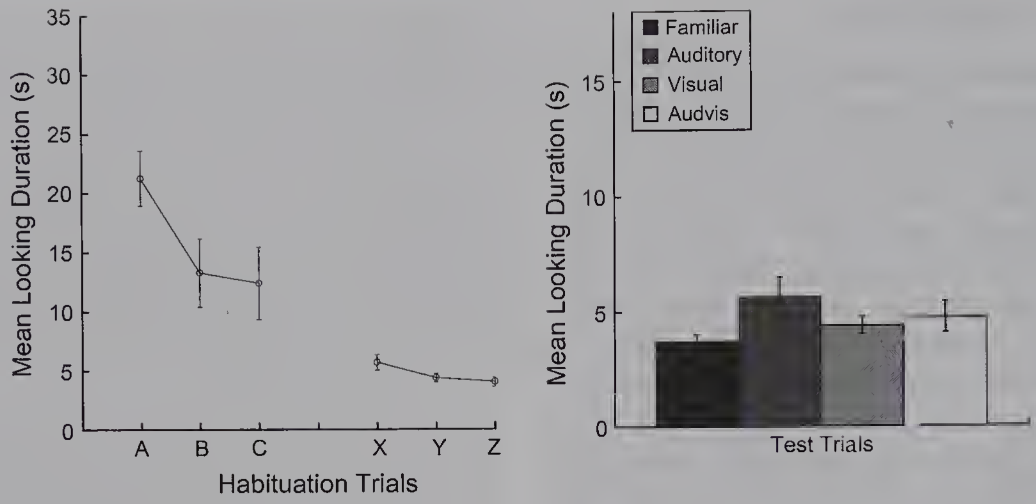


Figure 4. Results from Experiment 3.
Note. The left panel shows the duration of looking during the first three (A, B, and C) and last three (X, Y, and Z) habituation trials, and the right panel shows the duration of looking in the test trials. Error bars indicate the standard error of the mean.

reached conventional levels of statistical significance if sample size were larger. As indicated earlier, however, any legitimate comparison between the outcome in the current experiment and the outcomes in the previous experiments makes it imperative that sample size be held constant.

The results from this experiment indicate that motion plays an important role in infant perception and discrimination of dynamic audiovisual sequences. When motion cues were reduced, infants only responded to the auditory order changes. What makes this finding especially interesting is that the vast majority of studies of sequence learning in human and nonhuman adults (Keele et al., 2003; Terrace, 1993) have used spatially static stimulus arrays and have, nonetheless, obtained successful learning of complex sequences. This suggests that infants' apparent reliance on motion in dynamic audiovisual sequences may serve two functions. First, it may recruit greater attention to the event because moving visual stimuli are known to elicit greater attention than static stimuli (Shaddy & Colombo, 2004; Wilcox & Clayton, 1968). Second, the sequential emergence of the objects at the beginning of each iteration of the event and their subsequent common motion are likely to highlight the event's serial organization. As indicated earlier, infants of this age rely on common motion for object perception (Kellman & Spelke, 1983; Kellman et al., 1986), and thus, it is reasonable to conclude that they also rely on motion for perception of sequential structure.

Experiment 4

In a study of sequence learning with the mobile conjugate reinforcement procedure, Gulya et al. (1998) found that 3-month-old infants could learn to move a series of three mobiles in a particular sequence and that they remembered the order in which they moved them a day later. In a second study, using the same mobile conjugate reinforcement procedure, Gulya et al. (1999) found that 6-month-old infants could learn and remember the order in which they moved a series of five different mobiles and that they remembered this order a day later. Together, these results provide a rough estimate of the limits of sequence learning at these two ages. If these findings are extrapolated to the learning of the kinds of dynamic audiovisual sequences presented in the current study, it is reasonable to predict that 4-month-old infants should be able to perceive, learn, and discriminate sequences consisting of more than three elements.

Although this prediction is reasonable, it is complicated by the fact that young infants' response to the global sequential organization of audiovisually specified sequences can compete with their propensity to attend to local event attributes. That is, Lewkowicz (2004) found that 4-month-old infants do not respond to the overall organization of an audiovisual sequence unless the impact that each object making up a given sequence is blocked from view. To account for this finding, Lewkowicz advanced an attention allocation hypothesis. This hypothesis was based on findings that 4-month-old infants tend to attend to the impact part of moving/sounding objects (Spelke, Born, & Chu, 1983) and on the conclusion that this tendency causes them to miss the overall sequential organization of a dynamic sequence.

If 4-month-old infants' tendency to focus on the audiovisual impact part of an event can affect the allocation of attention to the sequential organization of a dynamic sequence, then other stimulus attributes also might affect it. For example, it is possible that increasing the number of sequence elements may actually shift attention away from the audiovisual impact part of the event and may enable infants to perceive an event's overall sequential organization. This may be because a greater number of sequence elements presented over the same amount of time may make it more difficult for infants to attend to the impact part of each object's motion simply because they go by more rapidly. Thus, although better learning of longer sequences might at first blush appear to be paradoxical, this result actually makes sense when event dynamics are considered. Indeed, this prediction is supported by findings from studies with adults showing that they tend to focus on the global properties of multiple-object scenes rather than on the properties of the individual objects (Navon, 1977, 2003; Pomerantz, 2003). For example, in a scene consisting of multiple gratings, adults perceive the mean orientation of the gratings rather than their individual orientations (Parkes, Lund, Angelucci, Solomon, & Morgan, 2001). Similarly, in a scene consisting of objects differing in size, adults tend to compute the mean size of the overall set of objects (Ariely, 2001; Sussman & Scholl, in press). Thus, the general principle appears to be that the greater the amount of information, the more attention is allocated to global stimulus features. This kind of global precedence effect also has been found in infants. In particular, studies have shown that 4-month-old but not younger infants tend to focus more on global than on local stimulus attributes (Colombo, Mitchell, Coldren, & Freeseman, 1991; Freeseman, Colombo, & Coldren, 1993). As a result, it is reasonable to conclude that by 4 months of age infants might attend more to the overall organization of a sequence when it consists of a greater number of elements and, in the process, may exhibit facilitation of learning.

If it is found that learning of the sequential organization of a longer dynamic audiovisual sequence is better than the learning of a shorter one, then this would suggest that, up to some limit the relationship between sequence complexity and learning is nonlinear in early infancy. Such a nonlinear effect is actually consistent with general principles of infant information processing (Cohen, Chaput, & Cashon, 2002). According to these principles, infants attempt to process events in terms of their highest information level possible and only drop to a lower information level when the processing system gets overloaded. In the case of dynamic audiovisual sequences, infants' first tendency may be to attempt to apprehend the sequential organization of the event. If, on the one hand, the event consists of a low number of objects moving relatively slowly, infants may be perceptually captured by the audiovisual impacts made by each object and, as a result, may drop to a lower level of processing. If, on the other hand, the event consists of a larger number of objects, then the global precedence effect may redirect their attention to the event's global attributes and they may respond to the event's overall sequential organization.

The purpose of Experiment 4 was to test the attention allocation hypothesis and the prediction that presentation of a longer sequence over the same amount of time may actually facilitate perception, learning, and discrimination. To do so, 4-month-old infants' response to four-element dynamic audiovisual sequences was investigated in this experiment. In addition, the purpose of Experiment 4 was to provide a strong test of the attention allocation hypothesis. This was done by allowing infants to see the impact part of the event. By doing so, it was possible to determine how the tendency to focus on the local (i.e., impact) part of the event and the competing tendency to attend to the global organization of the sequence would affect response. If the infants continue to focus on the event's local attributes, then, like the infants in Experiment 1 in Lewkowicz's (2004) study, they should only respond to the bisensory order change. In contrast, if they now shift their attention to the event's global attributes, then they should exhibit more robust evidence of the detection of serial order changes and, thus, should respond to the unisensory order changes as well.

Method

Participants. Twenty-four 4-month-old infants (*M* age = 19 weeks, *SD* = 0.75 weeks; 10 boys and 14 girls) contributed data in this experiment. One additional infant was tested but was not included because of fussing.

Apparatus and stimuli. All aspects of this experiment were the same as those of Experiment 1 with two important exceptions. First, the sequences consisted of four distinct objects and sounds. Three of the objects and sounds were identical to those presented in Experiment 1. The new fourth object resembled a star that had an orange interior and purple apices and was accompanied by a sound corresponding to a blip.wav file. Second, here the gray rectangle shown in Figure 1a was still visible except that now it was placed behind the objects' motion path. As a result, infants could now see the objects along their entire motion trajectory and, thus, could see and hear the impact that each object made as it changed its direction of motion.

Procedure. As in Experiment 1, each movie began with the appearance of the spout, ramp, and the gray rectangle. Thirty milliseconds later, the first object emerged from the spout and moved down. The next three objects then emerged 0.53, 1.03, and 1.53 s following the emergence of the first object, respectively. All four objects moved down at the same and constant speed and contacted the ramp 1.83, 2.33, 2.83, and 3.33 s following their emergence from the spout, respectively. Following contact with the ramp, the objects turned to the right, slid down the ramp, and came to a stop 4.53, 4.78, 5.15, and 5.43 s following their emergence from the spout, respectively. Once the last object came to a stop, all four objects remained visible and stationary for 0.5 s. The objects then disappeared, and 0.6 s later, the event started again and continued to be repeated for the duration of the trial. To ensure that infants were not basing their discrimination on initial element differences, the initial element of the contrasting sequences was kept constant. The overall design of this experiment can be seen in Table 2.

Results and Discussion

Inspection of the F test trial data indicated that 1 infant exhibited spontaneous regression to the mean in this trial; therefore, this infant's data were eliminated from all subsequent analyses. The habituation and test data from the remaining 23 infants can be seen in Figure 5, and as can be seen in the left panel of Figure 5, looking declined during the habituation phase. A mixed 2 × 6 (Habituation Group × Habituation Trials) ANOVA, with habituation group as the between-subjects factor and habituation trials as the within-subjects factor, indicated that infants exhibited a statistically reliable decrement in looking during the habituation phase, $F(5, 105) = 17.3, p < .001$, and that the nature of this decrement was not affected by the particular sequence presented during this phase.

The right panel of Figure 5 shows the mean duration of looking in the test trials. A mixed 2 × 4 (Habituation Group × Test Trial Type) ANOVA, with habituation group as the between-subjects factor and test trial type as the within-subjects factor, indicated that responsiveness differed across the test trials, $F(3, 63) = 2.88, p < .05$, but that it was not affected by the particular habituation sequence. Planned comparison tests indicated that infants exhibited significant response recovery in the A test trial, $F(1, 21) = 6.44, p < .025$; the V test trial, $F(1, 21) = 4.84, p < .05$; and the AV test trial, $F(1, 21) = 11.5, p < .01$.

The outcome of this experiment was consistent with the attention allocation hypothesis and with the prediction that infants may exhibit more robust detection of serial order changes. This was evident in the fact that infants successfully detected the unisensory as well as bisensory order changes. The finding that infants detected all three types of order changes despite the fact that they had access to the audiovisual impact part of the event provides strong support for the assumptions underlying the attention allocation hypothesis and its prediction that increasing the number of sequence elements should shift attention to the overall serial organization of the sequences. Had infants not shifted their attention to the event's global sequential organization, they would not have detected the unisensory order changes. Moreover, the finding that infants exhibited successful sequence perception and discrimination despite the fact that the initial elements of the contrasting sequences did not differ provides further proof that they perceived the events' global sequential organization and that primacy effects did not play a role in responsiveness.

General Discussion

The results from the current study replicate as well as extend the findings from Lewkowicz's (2004) study of infant perception and discrimination of dynamic audiovisual sequences. They do so in a number of ways. First, they show that infants as young as 3

Table 2
The Sequences Presented to Each Group in the Habituation and Test Trials in Experiment 4

Habituation group	Test trials Familiar (F)	 Auditory (A)	 Visual (V)	 Auditory – Visual (AV)
ABCD	Auditory-ABCD/ Visual-ABCD	Auditory-ADBC/ Visual-ABCD	Auditory-ABCD/ Visual-ADBC	Auditory-ADBC/ Visual-ADBC
ADBC	Auditory-ADBC/ Visual-ADBC	Auditory-ABCD/ Visual-ADBC	Auditory-ADBC/ Visual-ABCD	Auditory-ABCD/ Visual-ABCD

Note. The letters A, B, C, and D designate each object and its corresponding impact sound (see the Method section for details).

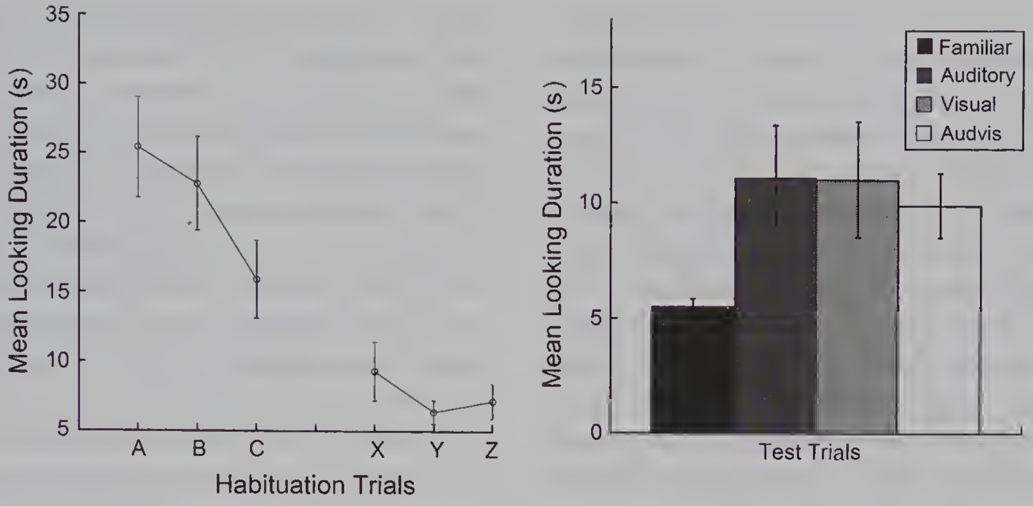


Figure 5. Results from Experiment 4.
Note. The left panel shows the duration of looking during the first three (A, B, and C) and last three (X, Y, and Z) habituation trials, and the right panel shows the duration of looking in the test trials. Error bars indicate the standard error of the mean.

months of age can perceive and discriminate three-element dynamic audiovisual sequences and that they, like 4-month-old infants, rely on multisensory redundancy for successful perception and discrimination (Experiment 1). Second, they show that 4-month-old infants can discriminate between contrasting dynamic audiovisual sequences even when their initial elements are held constant (Experiments 2 and 4). This indicates that primacy effects do not account for infants' successful detection of sequence differences. Third, they show that motion plays an important role in perception and discrimination of dynamic audiovisual sequences in that infants exhibited less robust responsiveness to audiovisual sequences consisting of looming rather than moving objects (Experiment 3). Finally, the current results show that 4-month-old infants can perceive and discriminate longer sequences in that they not only perceived and discriminated four-element sequences but also exhibited more robust responsiveness to such sequences (Experiment 4).

The finding that 3-month-old infants relied on multisensory redundancy to encode and detect serial order is consistent with evidence from human infant (Bahrick & Lickliter, 2000; Bahrick et al., 2004; Lewkowicz & Kraebel, 2004), human adult (Hershenson, 1962; Stein & Meredith, 1993; Summerfield, 1979), and animal adult (Partan & Marler, 1999; Rowe, 1999) studies indicating that attention, learning, and discrimination are substantially better when events are specified redundantly in multiple sensory modalities. At the same time, however, the findings from the other experiments that 4-month-old infants detected serial order changes in less complex three-element sequences (Experiment 2) and in more complex four-element sequences (Experiment 4) regardless of whether they were redundantly specified indicate that the value of multisensory redundancy in early infancy is modulated by the particular nature of the information to be processed and by task difficulty (Bahrick et al., 2004). That is, when the perceptual task of encoding serial order is made difficult by presenting fewer sequence elements, infants tend to rely on multisensory redundancy to extract the serial order information. In contrast, when the perceptual task is made easier by presenting a larger number of sequence elements and thus shifting infants' attention to the global event attributes, infants no longer need to rely on multisensory redundancy for perceiving and encoding the event's sequential organization.

The results from Experiments 2 and 4 clearly demonstrate that primacy effects do not account for young infants' response to dynamic audiovisual sequences. In particular, they indicate that infants do not base their detection of serial order changes on initial element differences. As a result, they support Lewkowicz's (2004) conclusion that the pattern of findings in that study did not reflect primacy effects.

The finding in Experiment 3 that infants only perceived the auditory order attributes of sequences composed of looming/sounding objects shows clearly that motion contributes to infant perception of sequential structure. Most likely, infants' attention is captured and/or recruited by moving objects to such an extent that it causes them to focus on the objects' serial ordering. Indeed, this beneficial effect of motion on perception has been found in a number of other studies of infant response to multisensory events. For example, when infants are given a choice between two static visual stimuli and a sound that corresponds to one of them in terms of onset and offset synchrony and temporal rate, they do not make intersensory matches (Lewkowicz, 1985). Similarly, when infants are habituated to a spatially static visual stimulus (i.e., a flashing checkerboard) and a pulsing tone, they respond primarily to the auditory information (Lewkowicz, 1988a, 1988b). In contrast, when infants are habituated to a bouncing visual stimulus and an accompanying impact sound, they respond to changes in the rate of visual and auditory information (Lewkowicz, 1992b) as well as to the temporal synchrony relation between these two sources of information (Lewkowicz, 1996). Likewise, when infants are given a choice between two moving visual stimuli and a sound that corresponds to the impact that one of these visual stimuli makes, they detect the temporal relationship between the visible and the audible impacts (Lewkowicz, 1992a). In other words, infants can flexibly change how much attention they devote to a particular event depending on the particular type of information provided. When objects move, infants devote substantial attentional resources to motion-related information and, in the process, learn to extract specific spatiotemporal information that they can then use for relating the objects' behaviors to what they are hearing. Perhaps the most dramatic illustration of this fact is evidence that infants as young as 6 months of age can take advantage of specific spatiotemporal relations between ambiguously moving objects and concurrently presented sounds and by doing so end up perceiving audiovisual illusions (Scheier et al., 2003).

The results from Experiment 4 extend the findings from Experiments 1–3 and the findings from Lewkowicz's (2004) study by demonstrating that 4-month-old infants can perceive longer, four-element dynamic audiovisual sequences. One particularly interesting aspect of the results from Experiment 4 is that infants exhibited robust detection of serial order changes despite the fact that the impact part of the event was visible. This contrasts with infants' failure to detect any serial order changes in their response to three-element sequences in Lewkowicz's (2004) Experiment 1 where the impact part of the event also was visible. At first blush, the results from Experiment 4 appear to be paradoxical. When considered in the context of the attention allocation hypothesis, however, they actually make a good deal of sense. According to this hypothesis, perception and discrimination of dynamic audiovisual sequences depend on the particular nature of sequences and the flexible allocation of attention to different aspects of such sequences. The results from the current study suggest that young infants' response to dynamic audiovisual sequences of different lengths/complexity depends on two competing perceptual processes. One compels young infants to attend to the event's local attributes (i.e., the audiovisual impact) and the other compels them to attend to the event's global attributes (i.e., the overall serial organization). Which one dominates processing depends on the length of the sequence. When the sequence consists of fewer elements, infants attend to the local attributes and, thus, find it more difficult to detect serial order changes. In contrast, when the sequence consists of more elements, infants attend to the global attributes and, thus, find it easier to detect serial order changes.

Although the attention allocation hypothesis and its a priori predictions received strong empirical support in the current study, future studies might examine eye movements to test the hypothesis even more directly. The straightforward prediction arising from the hypothesis is that in the case of three-element sequences whose audiovisual impact is not blocked from view, 4-month-old infants should fix their eye gaze more on the impact part of the event and, as a result, should engage in less tracking of the various objects. In contrast, in the case of four-element sequences whose audiovisual impact is also not blocked from view, 4-month-olds should track the motion of the objects more and, as a result, should fix their gaze less on the impact part of the event.

The current findings, together with those from Lewkowicz's (2004) study, clearly indicate that young infants are sensitive to information specifying the serial organization of dynamic audiovisual sequences. At the same time, however, this set of findings does not permit any firm conclusions regarding the specific type of information that mediated successful discrimination of serial order differences. This is because the contrasting sequences in the current and in Lewkowicz's (2004) study differed in terms of adjacent as well as ordinal position relations. A recently completed study in the author's laboratory (Lewkowicz & Berent, 2008) helps shed some light on what particular information infants respond to when processing the types of dynamic audiovisual sequences presented here. In this study, we habituated 4-month-old infants to several different four-element sequences consisting of the same types of objects and impact sounds presented here. Three of these objects and their sounds were presented in different ordinal positions across different habituation trials, whereas one, the target element, was presented in an invariant ordinal position. Results showed that infants detected a change in the ordinal position of the target element when that change occurred in the context of familiar sequence elements but that they did not detect the change when it occurred in the context of three novel sequence elements. Moreover, results indicated that infants' response to ordinal position differences in the familiar context was based on their detection of statistical relations (i.e., transitional probability relations among adjacent sequence elements) and not on their detection of ordinal position information. Based on these results, the most conservative interpretation of the data from the current study is that infants' responsiveness was probably based on their detection of statistical relations and not of ordinal relations.

Overall, the findings from the current study indicate that the ability to perceive dynamic audiovisual sequences emerges early in human development. This ability surely provides the essential developmental foundation for the subsequent emergence of higher level perceptual and cognitive skills. For example, it is known that 7-month-old infants can extract the syntactic structure of a linguistic utterance (Marcus et al., 1999), 8-month-old infants can produce a series of correctly sequenced actions to achieve a goal (Piaget, 1952), 11-month-olds can parse the temporal structure of ongoing behaviors along boundaries related to the initiation and completion of intentions (Baldwin, Baird, Saylor, & Clark, 2001), and 12-month-old infants understand that actions that occur in a sequence can imply an overarching goal (Woodward & Sommerville, 2000). Successful performance in all these tasks depends on infants' ability to perceive the underlying serial ordering of sequence components. The fact that the kinds of sequence perception skills investigated here are already present by 3 months of age suggests that the perceptual foundation for the development of the higher level skills emerges early in life.

The challenge for future studies is to determine when particular sequence perception, learning, and discrimination skills emerge; how they change and improve with development; and how they contribute to the development of higher level perceptual and cognitive skills. Currently, it is not clear what particular sequence perception skills emerge at what point in development and why they emerge at that particular time. Nelson (1986) and Greenfield (1991) have proposed that humans are endowed with a sensitivity to sequential structure at birth and that this sensitivity provides the developmental basis for the emergence of higher level event representation and comprehension skills. Although the finding that infants as young as 3 months of age can perceive and discriminate dynamic audiovisual sequences is generally consistent with these theoretical views, it does not mean that the ability to perceive the types of sequential event structures presented here is present at birth. First, motion perception is initially relatively poor, begins to emerge at 1 month of age, and continues to improve over the next several months. As a result, it is unlikely that newborn infants would be able to extract serial order information from an array of several moving objects. Second, despite the fact that infants come into this world with a general sensitivity to temporal information (Nazzi, Bertoncini, & Mehler, 1998), this sensitivity is relatively poor initially and improves over the 1st year of life (Morrongiello, Kulig, & Clifton, 1984; Trehub, Schneider, & Henderson, 1995). Finally, experience with moving/sounding objects during the first months of life is likely to contribute in important ways to the eventual acquisition of more sophisticated perceptual skills and, by extension, to the acquisition of sequence perception and learning skills.

The foregoing considerations suggest one reasonable preliminary theoretical model of the emergence of sequence perception skills. According to this model, an initial general sensitivity to temporal information provides infants with the first perceptual tools to begin to discover the hidden temporal structure inherent in multisensory events. As infants begin to perceptually differentiate this structure, they first discover the relatively simple statistical relations linking adjacent sequence elements. With further perceptual experience and differentiation, infants progressively discover increasingly more complex sequential relations. Thus, according to this scenario, following the discovery of statistical relations, they next discover that the elements of a sequence can be related to one another in terms of their ordinal position. Then, they discover that sequence elements can be related in a rule-bound manner (i.e., in terms of syntactic-like rules). Finally, they discover that sequence elements can be related in terms of hierarchical sequential relations (as in recursion). One assumption underlying this model is that the discovery of each type of sequential relation emerges from the earlier acquisition of the previous and presumably less complex sequential relation. Although this developmental model is admittedly speculative, it is consistent with currently available evidence on sequence perception and learning in infancy. Moreover, this model is developmental in nature and specifies the particular types of information that infants are able to perceive as they continue to improve their sequence perception abilities. Needless to say, it is up to future studies to determine whether the current model offers a reasonable account of the development of sequence perception, learning, and discrimination skills in infancy and whether it provides insights into the developmental basis for the acquisition of higher level cognitive skills during infancy and beyond.

References

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science, 12,* 157–162.

Bahrick, L. E., & Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Developmental Psychology, 36,* 190–201.

Bahrick, L. E., Lickliter, R., & Flom, R. (2004). Intersensory redundancy guides the development of selective attention, perception, and cognition in infancy. *Current Directions in Psychological Science, 13,* 99–102.

Baldwin, D. A., Baird, J. A., Saylor, M. M., & Clark, M. (2001). Infants parse dynamic action. *Child Development, 72,* 708–717.

Bauer, P. J., Hertsgaard, L. A., Dropik, P., & Daly, B. P. (1998). When even arbitrary order becomes important: Developments in reliable temporal sequencing of arbitrarily ordered events. *Memory, 6,* 165–198.

Bauer, P. J., Wiebe, S. A., Waters, J. M., & Bangston, S. K. (2001). Reexposure breeds recall: Effects of experience on 9-month-olds' ordered recall. *Journal of Experimental Child Psychology, 80,* 174–200.

Bertenthal, B. I., Proffitt, D. R., & Cutting, J. E. (1984). Infant sensitivity to figural coherence in biomechanical motions. *Journal of Experimental Child Psychology, 37,* 213–230.

Bertenthal, B. I., Proffitt, D. R., Kramer, S. J., & Spetner, N. B. (1987). Infants' encoding of kinetic displays varying in relative coherence. *Developmental Psychology, 23,* 171–178.

Canfield, R. L., & Haith, M. M. (1991). Young infants' visual expectations for symmetric and asymmetric stimulus sequences. *Developmental Psychology, 27,* 198–208.

Carver, L. J., & Bauer, P. J. (1999). When the event is more than the sum of its parts: 9-month-olds' long-term ordered recall. *Memory, 7,* 147–174.

Chen, S., Swartz, K. B., & Terrace, H. S. (1997). Knowledge of the ordinal position of list items in rhesus monkeys. *Psychological Science, 8,* 80–86.

Cohen, L. B., Chaput, H. H., & Cashon, C. H. (2002). A constructivist model of infant cognition. *Cognitive Development, 17,* 1323–1343.

Colombo, J., Mitchell, D. W., Coldren, J. T., & Freeseman, L. J. (1991). Individual differences in infant visual attention: Are short lookers faster processors or feature processors? *Child Development, 62,* 1247–1257.

Cornell, E. H., & Bergstrom, L. I. (1983). Serial-position effects in infants' recognition memory. *Memory & Cognition, 11,* 494–499.

Dannemiller, J. L., & Freedland, R. (1991). Detection of relative motion by human infants. *Developmental Psychology, 27,* 67–78.

Freeseman, L. J., Colombo, J., & Coldren, J. T. (1993). Individual differences in infant visual attention: Four-month-olds' discrimination and generalization of global and local stimulus properties. *Child Development, 64,* 1191–1203.

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science, 13,* 431–436.

Gómez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition, 70,* 109–135.

Gómez, R. L., & Maye, J. (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy, 7,* 183–206.

Greenfield, P. M. (1991). Language, tools and brain: The ontogeny and phylogeny of hierarchically organized sequential behavior. *Behavioral and Brain Sciences, 14,* 531–595.

Gulya, M., Rovee-Collier, C., Galluccio, L., & Wilk, A. (1998). Memory processing of a serial list by young infants. *Psychological Science, 9,* 303–307.

Gulya, M., Sweeney, B., & Rovee-Collier, C. (1999). Infants' memory processing of a serial list: List length effects. *Journal of Experimental Child Psychology, 73,* 72–91.

Hauser, M. D., Weiss, D., & Marcus, G. (2002). Rule learning by cotton-top tamarins. *Cognition, 86,* B15–B22.

Hershenson, M. (1962). Reaction time as a measure of intersensory facilitation. *Journal of Experimental Psychology, 63,* 289–293.

Jaffe, J., Beebe, B., Feldstein, S., Crown, C. L., & Jasnow, M. D. (2001). Rhythms of dialogue in infancy. *Monographs of the Society for Research in Child Development, 66*(Serial No. 265).

Jusczyk, P. W., Friederici, A. D., Wessels, J. M., Svenkerud, V. Y., & Jusczyk, A. M. (1993). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory & Language, 32,* 402–420.

Jusczyk, P. W., Luce, P. A., & Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory & Language, 33,* 630–645.

Kaufmann, F., Stucki, M., & Kaufmann-Hayoz, R. (1985). Development of infants' sensitivity for slow and rapid motions. *Infant Behavior & Development, 8,* 89–98.

Kaufmann-Hayoz, R., Kaufmann, F., & Stucki, M. (1986). Kinetic contours in infants' visual perception. *Child Development, 57,* 292–299.

Keele, S. W., Ivry, R., Mayr, U., Hazeltine, E., & Heuer, H. (2003). The cognitive and neural architecture of sequence representation. *Psychological Review, 110,* 316–339.

Kellman, P. J., & Arterberry, M. E. (1998). *The cradle of knowledge: Development of perception in infancy.* Cambridge, MA: MIT Press.

Kellman, P. J., & Spelke, E. S. (1983). Perception of partly occluded objects in infancy. *Cognitive Psychology, 15,* 483–524.

Kellman, P. J., Spelke, E. S., & Short, K. R. (1986). Infant perception of object unity from translatory motion in depth and vertical translation. *Child Development, 57,* 72–86.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition, 83,* B35–B42.

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior: The Hixon symposium* (pp. 123–147). New York: Wiley.

Lewkowicz, D. J. (1985). Bisensory response to temporal frequency in 4-month-old infants. *Developmental Psychology, 21*, 306–317.

Lewkowicz, D. J. (1988a). Sensory dominance in infants: I. Six-month-old infants' response to auditory-visual compounds. *Developmental Psychology, 24*, 155–171.

Lewkowicz, D. J. (1988b). Sensory dominance in infants: II. Ten-month-old infants' response to auditory-visual compounds. *Developmental Psychology, 24*, 172–182.

Lewkowicz, D. J. (1989). The role of temporal factors in infant behavior and development. In I. Levin & D. Zakay (Eds.), *Time and human cognition: A life-span perspective* (pp. 9–62). Amsterdam: North-Holland.

Lewkowicz, D. J. (1992a). Infants' response to temporally based intersensory equivalence: The effect of synchronous sounds on visual preferences for moving stimuli. *Infant Behavior & Development, 15*, 297–324.

Lewkowicz, D. J. (1992b). Infants' responsiveness to the auditory and visual attributes of a sounding/moving stimulus. *Perception & Psychophysics, 52*, 519–528.

Lewkowicz, D. J. (1996). Perception of auditory-visual temporal synchrony in human infants. *Journal of Experimental Psychology: Human Perception & Performance, 22*, 1094–1106.

Lewkowicz, D. J. (2000). The development of intersensory temporal perception: An epigenetic systems/limitations view. *Psychological Bulletin, 126*, 281–308.

Lewkowicz, D. J. (2004). Perception of serial order in infants. *Developmental Science, 7*, 175–184.

Lewkowicz, D. J., & Berent, I. (2008). *Sequence learning in 4 month-old infants: Do infants represent ordinal information?* Poster presented at the International Conference on Infant Studies.

Lewkowicz, D. J., & Kraebel, K. (2004). The value of multimodal redundancy in the development of intersensory perception. In G. Calvert, C. Spence, & B. Stein (Eds.), *Handbook of multisensory processing*. Cambridge, MA: MIT Press.

Marcus, G., Vijayan, S., Rao, S., & Vishton, P. (1999). Rule learning by seven-month-old infants. *Science, 283*, 77–80.

Martin, J. G. (1972). Rhythmic (hierarchical) versus serial structure in speech and other behavior. *Psychological Review, 79*, 487–509.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*, 229–239.

Merriman, J., Rovee-Collier, C., & Wilk, A. (1997). Exemplar spacing and infants' memory for category information. *Infant Behavior & Development, 20*, 219–232.

Morrongiello, B. A., Kulig, J. W., & Clifton, R. K. (1984). Developmental changes in auditory temporal perception. *Child Development, 55*, 461–471.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology, 9*, 353–383.

Navon, D. (2003). What does a compound letter tell the psychologist's mind? *Acta Psychologica, 114*, 273–309.

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception & Performance, 24*, 756–766.

Neil, P. A., Chee-Ruiter, C., Scheier, C., Lewkowicz, D. J., & Shimojo, S. (2006). Development of multisensory spatial integration and perception in humans. *Developmental Science, 9*, 454–464.

Nelson, K. (1986). *Event knowledge: Structure and function in development*. Hillsdale, NJ: Erlbaum.

Orlov, T., Yakovlev, V., Hochstein, S., & Zohary, E. (2000). Macaque monkeys categorize images by their ordinal number. *Nature, 404*, 77–80.

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience, 4*, 739–744.

Partan, S., & Marler, P. (1999). Communication goes multimodal. *Science, 283*, 1272–1273.

Piaget, J. (1952). *The origins of intelligence in children*. New York: International Universities Press.

Pomerantz, J. R. (2003). Wholes, holes, and basic features in vision. *Trends in Cognitive Sciences, 7*, 471–473.

Rowe, C. (1999). Receiver psychology and the evolution of multicomponent signals. *Animal Behaviour, 58*, 921–931.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1926–1928.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*, 27–52.

Scheier, C., Lewkowicz, D. J., & Shimojo, S. (2003). Sound induces perceptual reorganization of an ambiguous motion display in human infants. *Developmental Science, 6*, 233–244.

Shaddy, D. J., & Colombo, J. (2004). Developmental changes in infant attention to dynamic and static stimuli. *Infancy, 5*, 355–365.

Shea, S. L., & Aslin, R. N. (1990). Oculomotor responses to step-ramp targets by young human infants. *Vision Res, 30*, 1077–1092.

Smith, P., Loboschefski, T. W., Davidson, B. K., & Dixon, W. E. (1997). Scripts and checkerboards: The influence of ordered visual information on remembering locations in infancy. *Infant Behavior & Development, 20*, 549–552.

Spelke, E. S., Born, W. S., & Chu, F. (1983). Perception of moving, sounding objects by four-month-old infants. *Perception, 12*, 719–732.

Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. Cambridge, MA: MIT Press.

Summerfield, A. Q. (1979). Use of visual information in phonetic perception. *Phonetica, 36*, 314–331.

Sussman, R. S., & Scholl, B. J. (in press). The flexibility of statistical summary representations in vision. *Perception & Psychophysics.*

Terrace, H. S. (1993). The phylogeny and ontogeny of serial memory: List learning by pigeons and monkeys. *Psychological Science, 4,* 162–169.

Terrace, H. S., Son, L. K., & Brannon, E. M. (2003). Serial expertise of rhesus macaques. *Psychological Science, 14,* 66–73.

Trehub, S. E., Schneider, B. A., & Henderson, J. L. (1995). Gap detection in infants, children, and adults. *Journal of the Acoustical Society of America, 98*(Pt. 1), 2532–2541.

von Hofsten, C. (1980). Predictive reaching for moving objects by human infants. *Journal of Experimental Child Psychology, 30,* 369–382.

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin, 88,* 638–667.

Wilcox, B. M., & Clayton, F. L. (1968). Infant visual fixation on motion pictures of the human face. *Journal of Experimental Child Psychology, 6,* 22–32.

Woodward, A. L., & Sommerville, J. A. (2000). Twelve-month-old infants interpret action in context. *Psychological Science, 11,* 73–77.

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin, 127,* 3–21.

Child Development, September/October 2008, Volume 79, Number 5, Pages 1555–1569

Longitudinal Analysis of Intrinsic Motivation and Competence Beliefs: Is There a Relation Over Time?

Birgit Spinath and Ricarda Steinmayr
University of Heidelberg

The present study explored whether competence beliefs and intrinsic motivation for different school domains show reciprocal effects over time. A sample of 670 German elementary school pupils ($M = 8.8$ years, $SD = 0.51$) was followed over 1 year. At 4 measurement occasions, children completed self-reports on their intrinsic motivation and competence beliefs for math, German, and school in general. Latent growth models revealed that intrinsic motivation and competence beliefs decreased over time. Comparing correlational and cross-lagged structural equation models yielded only weak evidence for cross-lagged influences between the 2 constructs. Results suggest that the developmental curves of competence beliefs and intrinsic motivation might be less inextricably interwoven than frequently assumed.

Maybe more than anything else, to be well equipped for lifelong learning, individuals need high, sustainable motivation to learn. In sharp contrast to this need for lifelong learning stands the observation that important prerequisites for learning, such as intrinsic motivation for school-related learning, diminish both during an individual's development (e.g., Gottfried, 1990; Gottfried, Fleming, & Gottfried, 2001; Spinath & Spinath, 2005b) and from one generation of pupils to the next (Cocodia et al., 2003; Howard, 2001). Therefore, there is a vital interest among researchers and educators to understand why intrinsic motivation for learning weakens and to find ways to preserve or reactivate the initially high intrinsic motivation for learning among young children. The present study investigated whether the well-established decline of competence beliefs and intrinsic motivation for learning during the 1st school years (i.e., in 6- to 10-year-old children) are not only parallel developments but mutually influencing each other. Specifically, we investigated the assumption that intrinsic motivation to learn declines as a consequence of children developing more realistic and therefore less positive ability self-concepts.

We chose this particular age for investigation because of the changes in children's concepts about ability between 6 and 10 years of age (e.g., Nicholls, 1978; Nicholls & Miller, 1984) and the co-occurring changes in the school environment (i.e., increasingly normative feedback). These developments should make potential reciprocal effects between competence beliefs and intrinsic motivation especially well observable. As it is common practice in most German elementary schools, children received their first grades during the investigated time. Usually, up until third grade only verbal evaluations are given that allow for less social comparisons among pupils. Therefore, if there is any relation between competence beliefs and intrinsic values over time, it should be very pronounced at this phase during school trajectory.

This study sets out to overcome some of the methodological short comings of prior research on related issues, in that its design follows established guidelines for investigating reciprocal effects in general (Marsh, Byrne, & Yeung, 1999) and relations between competence beliefs and intrinsic values specifically (Eccles, 2005). We realized a 1-year longitudinal assessment of competence beliefs and intrinsic motivation in three domains (math, German, and school in general) with four measurement occasions. Using structural equation models, it was tested whether cross-lagged paths between competence beliefs and intrinsic motivation at different measurement occasions were necessary to describe the data adequately (cross-lagged models) or whether more parsimonious correlational models provided equally good descriptions. Only if competence beliefs and intrinsic values were to show substantial cross-relation over time, would there be reason to assume that intrinsic motivation might drop as a consequence of less positive ability self-perceptions.

Correspondence concerning this article should be addressed to Birgit Spinath, Department of Psychology, University of Heidelberg, Hauptstrasse 47-51, D-69117 Heidelberg, Germany. Electronic mail may be sent to birgit.spinath@psychologie.uni-heidelberg.de.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7905-0022

Defining the Concepts of Intrinsic Motivation and Competence Beliefs

The concept of intrinsic motivation plays a vital role in different motivation theories (e.g., Deci & Ryan, 1985; Wigfield & Eccles, 2000). In the present article, we focus on the concept of intrinsic task values as included in the expectance-value theory of motivation by Eccles and her associates (cf. Wigfield & Eccles, 2000). Intrinsic task values denote the degree of positive affective evaluation of an activity, that is, liking and enjoyment, for reasons that lie within the activity itself rather than its consequences. Although intrinsic task values are not the only reason for learning, task enjoyment can be considered as the most desirable state for learners because learning comes as a by-product of engaging in a pleasurable activity. Moreover, it has been shown that intrinsic task values are the most important reasons for task engagement in elementary school children who do not yet differentiate other task values such as utility or importance (e.g., Eccles, Wigfield, Harold, & Blumenfeld, 1993; Wigfield & Eccles, 1992).

The second motivational construct we focus on in the present work is the concept of competence beliefs. In the expectancy-value model by Eccles and her colleagues (Eccles et al., 1983; Wigfield & Eccles, 2000), task values and expectations of future success are the two most proximal determinants of achievement-related choices. Expectancies for future success are usually operationalized by domain-specific ability self-perceptions. In school contexts, students are asked how good they think they are at certain tasks or in certain domains (e.g., Wigfield et al., 1997). It has been shown that children distinguish very early between expectancies and values for certain school domains (e.g., Eccles et al., 1993), so that in elementary school, children's ability self-perceptions and intrinsic values are distinct characteristics.

Why Should Intrinsic Motivation and Competence Beliefs Be Related?

The assumption that intrinsic motivation for learning and perceptions of one's competences are related is derived from both empirical observations and motivation theories. Experienced teachers as well as developmental researchers observe that not only intrinsic motivation for learning declines from elementary school age through adolescence but that the same is true for children's competence beliefs (e.g., Bouffard, Marcoux, Vezeau, & Bordeleau, 2003; Gottfried et al., 2001; Spinath & Spinath, 2005b; Wigfield et al., 1997). Whereas preschool children view their own competences in an overoptimistic way (cf. Stipek & MacIver, 1989), older children perceive themselves more realistically and sometimes even develop overpessimistic self-perceptions (e.g., R. S. Newman, 1984; Spinath & Spinath, 2005a; Stipek & Daniels, 1988; Wigfield et al., 1997). This parallel decline nourishes the plausible assumption that children's competence beliefs and their intrinsic values are causally linked: When children believe less in their competences, they are bound to lose their enjoyment of task engagement. Yet, whether the decline of competence beliefs and intrinsic motivation at this specific age are actually mutually influencing or only parallel but independent developments still needs to be discovered.

Reasons for declining competence beliefs. What are the reasons for children between 6 and 10 years old to perceive their competencies more realistically and therefore, in most cases, less positively? During this time, children become increasingly able to differentiate between effort and ability as causes of achievement outcomes and attain an understanding of ability as a capacity of the person (Nicholls, 1978; Nicholls & Miller, 1984). Once children attain this understanding about ability, they are more likely to use social comparison information to evaluate their own ability (e.g., Ruble, Boggiano, Feldman, & Loebl, 1980; Ruble, Parsons, & Ross, 1976). In addition, children learn to differentiate more strongly between ability domains such as physical, social, and academic abilities (Harter, 1982). Taken together, these developments contribute to children's growing competence to perceive their own ability accurately, that is, more in line with valid criteria of their actual ability (e.g., Harter, 1982; R. S. Newman, 1984; Spinath & Spinath, 2005a). Coming from overoptimistic self-perceptions, for most children, these changes mean a downward correction of their ability self-perceptions.

It is probably no coincidence that in most school systems, children are confronted with more social comparisons and normative feedback at a time when they develop a normative concept of ability. For example, in most German schools, children receive their first grades when they are about 9 years old, that is, in Grade 3. Up to that point, they would have had only verbal evaluations that are meant to focus on children's individual development rather than social comparisons among pupils. Thus, the age group of 9-year-olds that we investigated is at a special phase in their school trajectory where normative evaluations of competence become extremely salient (e.g., Eccles et al., 1993). We reasoned that if there are mutual influences between competence beliefs and intrinsic values over time, they should be very pronounced at this phase during school trajectory.

Theoretical grounds on which to assume a causal link between ability beliefs and intrinsic motivation. A theoretical background for the assumption that intrinsic motivation depends on competence perceptions is provided by White's (1959) theorizing about effectance motivation. According to White, individuals have an innate desire to feel competent. This effectance motivation drives individuals to engage in tasks in which mastery is at stake. Tasks that generate feelings of efficacy should be experienced as enjoyable or intrinsically rewarding and, as a by-product, prompt learning. Thus, in White's theory, feeling competent and being intrinsically motivated are the same. One has to note, however, that White's concept of feelings of efficacy or competence is not the same as the ability beliefs we focus on in the present investigation. Competence beliefs as operationalized by ability self-perceptions are cognitive representations about the level of ability one perceives to have. Nevertheless, White's theorizing provided the basis on which many motivation researchers grounded the assumption of a causal relation between competence beliefs and intrinsic motivation.

Building up on White's (1959) more general theorizing, Harter (1981a) refined and extended White's assumptions in an effort to provide testable hypotheses under a developmental perspective on competence beliefs. Harter distinguishes between competence beliefs as cognitive representations of the level of one's ability and motivational orientations (i.e., intrinsic and extrinsic motivation). In Harter's (1981a, p. 38) effectance motivation model, intrinsic pleasure in task engagement is a function of successful mastery of challenging tasks and perceived competence. Conversely, failure and perceived lack of competence should result in anxiety in mastery situations, an emotion known to be incompatible with enjoyable feelings. In postulating this sequential process, Harter disentangles perceived competence and motivational orientation and predicts that higher levels of perceived competence entail higher levels of intrinsic motivation. Some of the most prolific modern motivation theories share the assumption that more positive ability self-perceptions should generate more intrinsic motivation for a given task (e.g., Deci & Ryan, 1985; Wigfield & Eccles, 2000).

It is important to note that all theoretical reasoning about the relation between competence beliefs and intrinsic motivation assumes that prior competence beliefs influence subsequent intrinsic motivation. Although one could argue that prior intrinsic motivation could also have an influence on subsequent competence beliefs, there is no elaborated theoretical background making a strong point for such directional effects. Therefore, in the present investigation, we focus on the question whether there is evidence that competence beliefs might influence intrinsic motivation.

Why could competence beliefs and intrinsic motivation develop in parallel but without mutual influences? Although there is strong theoretical reasoning on which to ground the assumption that competence beliefs are important for intrinsic motivation, it is also worthwhile to look for theoretical reasons why such a link might not exist or be only weak. One could argue that it is not the absolute level of self-perceived ability that is important for the degree of task enjoyment but other kinds of ability-related self-perceptions. One line of reasoning can be drawn from Harter's (1981a) specifications to White's (1959) theory of effectance motivation. This specification concerns the relation between success and feelings of efficacy. Harter pointed out that feelings of efficacy, and therefore intrinsic motivation for task engagement, are maximized not by the mere quantity of success but by succeeding at optimally challenging tasks. In other words, intrinsic motivation should depend on an optimal fit between task difficulty and own competence. Such an optimal fit can be given at different levels of competence, so that not only the most competent individuals experience intrinsic motivation. Therefore, the degree of intrinsic motivation might not so much depend on the absolute level of normatively based ability perceptions but rather on perceiving oneself as successful at tasks that fit the level of one's competence. If this was true, we would expect no or only weak indications of normatively based ability self-perceptions influencing intrinsic motivation.

Taken together, developmentally focused motivational theories often assume a causal influence of competence beliefs on intrinsic motivation. Because competence beliefs undergo specific changes in the early school years that lead most children to less positive ability self-perception, declines in intrinsic motivation seem to be an inevitable consequence from this point of view. If, however, it could be shown that the development of intrinsic motivation is not or only weakly dependent on the development of competence beliefs, then the trajectories of the two might be disentangled by means of interventions.

Methodological Guidelines for Investigating Reciprocal Effects Over Time

After theorizing about the nature of the relation between competence beliefs and intrinsic motivation, we take a look at the empirical literature. There is ample evidence on the positive correlation among intrinsic motivation and competence beliefs within domains (e.g., Deci & Ryan, 1985; Harter, 1981b; Wigfield, 1994; Wigfield et al., 1997). Most of these studies are cross-sectional and correlational and allow no conclusions about a potentially causal relationship between the two constructs over time. In order to investigate whether prior competence beliefs might be influential for subsequent intrinsic motivation, longitudinal designs are needed.

In an effort to develop stronger methodological approaches to test for reciprocal effects between concepts in longitudinal designs, Marsh (1990) and Marsh et al. (1999, p. 156) provided valuable guidelines. These authors state that in order to detect causal relations, first, two constructs need to be shown to have a significant statistical relation (i.e., substantial path coefficients in structural equation models). Second, clear time precedence needs to be established in longitudinal studies with at least two and preferably more measurement occasions. Third, theoretical models must be tested by means of statistical techniques such as structural equation modeling (SEM) with all latent constructs inferred on the basis of multiple indicators. The most adequate description of the empirical data is derived by fitting the data to different theoretically plausible models among which to choose the best fitting one.

The present study realizes all these recommendations to examine the relationship between competence beliefs and intrinsic motivation. We are not aware of another study that conducted a 1-year longitudinal assessment of competence beliefs and intrinsic motivation in three domains (math, German, and school in general) with four measurement occasions. The two constructs were measured on a latent basis by three manifest indicators each. This design allows for comparing different theoretical models that differ in whether or not they assume causal relations between competence beliefs and intrinsic motivation. Specifically, it was tested whether cross-lagged paths between competence beliefs and intrinsic motivation at different measurement occasions were necessary to describe the data adequately (cross-lagged models) or whether more parsimonious correlational models provided equally good descriptions. According to a basic scientific principle, in the case of two competing models, which describe a phenomenon equally well, the more parsimonious model or theory should be preferred over the more complex one. Although the preference of the more parsimonious model in the case of an equal fit to the data does not rule out the existence of cross-lagged effects, such effects would be rather weak.

Of course, longitudinal designs cannot provide unequivocal evidence for causal influences. They are, however, very useful to test for necessary prerequisites of causal influences and give an impression of how strong such influences might be. Concerning the present investigation, this means that in case of no cross-lagged relations between prior competence beliefs and later intrinsic motivation, there is no basis for further investigating causal influences. Furthermore, the present design provides an estimation of the upper limit of the strength of potential causal influences. If there are only weak relations over time between competence beliefs and intrinsic motivation, this indicates that causal influences between the two might at best be weak.

Previous Longitudinal Studies Investigating the Link Between Intrinsic Motivation and Competence Beliefs

We identified five longitudinal studies investigating the link between intrinsic values for school-related learning and competence beliefs over time. Not all these studies followed the guidelines by Marsh and colleagues.

One of the most extensive studies concerning the link between intrinsic task values and competence beliefs (Jacobs, Lanza, Osgood, Eccles, & Wigfield, 2002) relied on the sample of the Michigan Childhood and Beyond Longitudinal Project initiated by Eccles and her associates. The development of children's self-perceptions, task values, and activity choices was investigated from Grade 1 to Grade 12. Hierarchical linear modeling was used to analyze the development of task values and self-concepts. When the development in task values was controlled for competence beliefs, the linear trends of the developmental curves were reduced in all investigated domains. Even though these results document the relation between ability beliefs and task values over time, they cannot be interpreted causally: Neither was the change in task values controlled for prior task values nor can the direction of a potential influence be clarified.

Nurmi and Aunola (2005) assessed intrinsic task values and ability self-beliefs in 6- to 7-year-olds at two measurement occasions during Grade 1 and Grade 2. By means of cluster-by-states analysis for longitudinal data, these authors found no association between academic self-concept and mathematical or reading task values when controlling for prior task values. The fact that this study found no evidence for potentially causal influences between competence beliefs and intrinsic values might be due to the young age of the investigated children who had probably not yet attained a concept of ability as a capacity of the person (Nicholls, 1978; Nicholls & Miller, 1984).

Whereas the two studies cited above did not use SEM, the following three studies used methodological approaches partly in line with the recommendations by Marsh and his colleagues. Skaalvik and Valas (1999) investigated mathematical and verbal ability self-concepts and motivation in three age groups (third, sixth, and eighth graders) over two school years with two measurement occasions. Motivation was measured by items reflecting either interest or the will to invest in a certain subject. Six SEMs were tested (two for each domain in each grade), resulting in little evidence for directional effects between competence beliefs and motivation. Specifically, with one exception, none of the cross-paths reached statistical significance (the path from motivation at Measurement Occasion 1 to ability self-concept at Measurement Occasion 2 in math was significant).

One feature of this study that might have impeded finding reciprocal effects is that the measurement occasions were rather far apart. The confirmation of directional effects might depend on the right time lag between measurement occasions. It has been argued that intervals between measurements should not be too long (e.g., Eccles, 2005) to prevent potential reciprocal influences from being masked by other processes. On the other hand, observations need to cover a sufficiently long time for effects to manifest themselves. To investigate both short-term and long-term reciprocal effects, the present investigation employed 3-month intervals and additionally tests for cumulative effects over the timespan of 1 year.

Another study that failed to find evidence for reciprocal effects between intrinsic motivation and competence beliefs is reported by Spinath and Spinath (2005b). Five cohorts of elementary school children ranging from Grade 1 to Grade 4 participated in the study. The cohorts were followed over a period of 2 years and children gave self-reports on their motivation (i.e., intrinsic values and learning goals) and their competence beliefs for school in general every 6 months. Data were analyzed by comparing a more complex reciprocal effects SEM with a more parsimonious correlational model omitting the reciprocal effects (method described in more detail earlier). There was no evidence of directional influences over time because the correlational model described the data as good as the reciprocal model.

Again, there are reasons why the study by Spinath and Spinath (2005b) might not have detected reciprocal effects (see Eccles, 2005). A first reason is that the different constructs were not measured as latent variables. Hence, models were not controlled for measurement errors and this might have masked true effects. A second reason is that the interval between measurement occasions might have been too long. As a third reason, it might be argued that intrinsic motivation and competence beliefs need to be measured domain specifically to find reciprocal influences. The present investigation reacts to all these arguments and measures the constructs on a latent basis, employs shorter measurement intervals, and investigates two domains (math and German) as well as school in general.

The only study using SEM that found evidence for reciprocal relations between an indicator of intrinsic motivation and competence beliefs was reported by Marsh, Trautwein, Lüdtke, Köller, and Baumert (2005). Two large samples of seventh graders gave self-reports on mathematical ability self-concept and interest twice within one school year. These authors found significant reciprocal standardized path weights (ranging from .04 to .10) between ability self-concept and interest. Although this methodologically sophisticated study produced evidence in favor of a reciprocal relationship over time, the effects were rather small and ran in both directions; that is, prior competence beliefs had effects on subsequent interest and vice versa.

One difference that might account for the inconsistent findings of the studies by Marsh et al. (2005) and Spinath and Spinath (2005b) is that the first assessed domain-specific competence beliefs and intrinsic motivation, whereas the latter focused on school in general. As mentioned before, assessing intrinsic values and competence beliefs on a more general level might mask reciprocal effects. In order to investigate whether there are differences between domains as well as between more and less specific measures, the present study compares domain-specific and domain-general intrinsic values and competence beliefs.

Taken together, previous studies suggest a need for stronger evidence supporting claims about the reciprocal effects between measures of intrinsic motivation and competence beliefs over time. Among the aforementioned studies, only three are largely, although not in every detail, in line with the guidelines proposed for the investigation of reciprocal relations in longitudinal studies (Marsh et al., 1999). Moreover, the studies cited above differ in several respects, ranging from different investigated age groups, domains, and indicators of intrinsic motivation to the use of different methodological approaches. Such differences might account for inconsistencies in empirical results.

The Present Investigation

The present study investigated whether intrinsic motivation and competence beliefs show reciprocal influences over time. Specifically, we investigated the assumption that intrinsic motivation to learn declines as a consequence of children's increasingly realistic and therefore less positive ability self-concepts. An answer to this question is given by addressing five individual aspects:

1. Do intrinsic values and competences beliefs decline during the investigated timespan? Such declines need to be investigated both in terms of an overall decline of the means of intrinsic values and competence beliefs in the sample as well as interindividual differences in these developments.
2. How stable are intrinsic values and competence beliefs over time? This is an important question because the more stable the two constructs are, the smaller the chance to find cross-lagged effects over time between them.
3. How strong are the concurrent correlations between the two concepts? The frequently observed medium to strong correlations between intrinsic values and competence beliefs gave rise to the assumption of causal influences between the two constructs.
4. Are prior competence beliefs and subsequent intrinsic values related over time when controlling for prior intrinsic values? And, vice versa, are prior intrinsic values and subsequent competence beliefs related when controlling for prior competence beliefs? Finding such relations over time would be in line with the assumption of causal influences between competence beliefs and intrinsic motivation.
5. Does a theoretical model that includes cross-lagged paths between intrinsic motivation and competence beliefs describe the empirical data better than a more parsimonious model with correlational paths but without cross-paths? This comparison of the two competing models is a second test of the assumption that intrinsic values and competence beliefs might share reciprocal, potentially causal effects over time.

Method

Participants

In 2002 and 2003, 22 third-grade classes from 13 German elementary schools in and around a midsized town were included in the investigation comprising a sample of 670 students. At the time of the first measurement occasion (June 2002), children were on average 8.8 years old ($SD = 0.51$). The investigation lasted 1 year so that children were 1 year older at the last measurement (May 2003). At the first measurement occasion, children were at the end of second grade, and at the last measurement occasion, children approached the end of third grade. The sample was roughly balanced for gender, with 297 (44.3%) of the children being boys. Most of the students were Caucasian (more than 98%) with German citizenship (86%). As typical for an elementary school student population of that cohort and region, children with Turkish citizenship (8%) represented the largest minority group.

Measures

Intrinsic values. Intrinsic values for math, German, and school in general were assessed by means of three items each. Children were asked to indicate on a 5-point scale ranging form *very much* (1) to *not at all* (5) how much they liked each of three different activities or school in general. The items read: "How much do you like mental arithmetics/math problems/doing calculations?" (math); "How much do you like reading/writing (stories, letters)/practicing spelling?" (German); "How much do you like the things you do in school/going to school/How much fun is school to you?" (school in general). For a better intuitive understanding of the scales, we recoded all items so that in the following higher scores denote higher intrinsic values.

A confirmatory factor analysis of the three correlated intrinsic value factors, conducted via SEM, yielded a good model fit at all four measurement occasions (comparative fit index [CFI] = 0.96 to 0.97; root mean square error of approximation [RMSEA] = .06 at all measurement occasions). Standardized path coefficients from the latent intrinsic value factors to the observed variables ranged from .52 to .90 (school in general), .61 to .81 (math), and .45 to .69 (German). Averaged over the four measurement occasions, the reliabilities of the intrinsic value scales were αs = .79, .73, and .53 (school in general, math, German). Because in math and in German, the items represent different activities that constitute each domain, the low internal consistency for German might represent the heterogeneity of this domain rather than a weakness of the scale. For heterogeneous domains, the alpha coefficient is not the appropriate reliability measure and other indicators, such as test–retest reliability, should also be considered. As depicted in

Table 4, test–retest reliabilities of the intrinsic values scale for German are higher than the internal consistencies and are at least satisfactory ($r_{tt} > .63$).

Competence beliefs. Children's competence beliefs for math, German, and school in general were assessed by means of three items each. The items were answered on 5-point scales and read: "How good are you at math/German/school?" with response choice ranging from *very good* (1) to *very bad* (5); "How easy is it for you to learn new things in math/German/school?" from *very easy* (1) to *very hard* (5); and "To which group of pupils do you belong in German/math/school?" from *the best* (1) to *the worst* (5). Again, items were recoded so that in the following higher scores denote higher competence beliefs.

A confirmatory factor analysis of the three correlated competence belief factors yielded a good model fit at all four measurement occasions (CFI = 0.98 to 0.99; RMSEA = .05 to .06). Standardized path coefficients from the latent competence belief factors to the observed variables ranged from .47 to .83 (school in general), .72 to .91 (math), and .67 to .85 (German). Internal consistencies for the three scales assessing competence beliefs, averaged over all four measurement occasions, were αs = .72, .87, and .83 (school in general, math, German).

Procedures

There was a total of four measurement occasions with an interval of 3 months between each test administration. Pupils were tested during a regular class in their classrooms. The testing took about 30 min and was conducted by teachers who did not teach in the tested class. Teachers had been trained to administer the items. Each item was read aloud to harmonize students' working speed.

Statistical Analyses

Missing data. As in all longitudinal studies, sample attrition needs to be taken care of. In the present study, missing data occurred when children missed class the day the testing took place. The main reason for absence was illness. We are not aware of any other reasons for children to miss the testing that might be systematically related to the testing and investigated variables. At each measurement occasion, approximately 80 students did not participate. Table 1 gives information about how many students participated at each measurement occasion. A total of 670 students participated at least at two measurement occasions. Complete data sets for all four measurements were available for 362 students, 270 students participated at three measurement occasions, and 38 participated at two measurement occasions. None of the children missed three measurement occasions. We tried to further explore the nature of the sample depletion by comparing children with missing measurements with the sample of complete data sets. We found no systematic differences in any investigated variable, age, or sex. Furthermore, there were only small amounts of missing data for individual items (less than 1%).

Instead of concentrating on the subsample that completed all measurement occasions, we accounted for missing data by means of full information maximum likelihood (FIML) estimations. We did so for two reasons. First, to maximize statistical power, it is desirable to include as many cases as possible in the analyses. Second, in longitudinal studies, it is difficult to show that data are missing at random and not in any way systematically. Studies indicate that even if data are not missing at random, FIML is still a more effective way of handling missing data than traditional procedures, such as listwise deletion, single- or

Table 1
Means, Standard Deviations, Subsample Without Missing Data (n), and Reliabilities (α) for Intrinsic Values and Competence Beliefs Scales in Three Domains for All Four Measurement Occasions

	t1			t2			t3			t4		
	M (SD)	*n*	α	*M (SD)*	*n*	α	*M (SD)*	*n*	α	*M (SD)*	*n*	α
Intrinsic values												
General	4.37 (0.77)	589	.75	4.29 (0.75)	586	.76	4.29 (0.79)	574	.80	4.23 (0.83)	576	.84
Math	3.89 (1.00)	590	.70	3.76 (0.99)	585	.71	3.79 (1.01)	576	.73	3.81 (1.05)	573	.76
German	4.13 (0.81)	590	.52	4.04 (0.85)	592	.60	4.11 (0.80)	574	.58	4.13 (0.74)	575	.43
Competence beliefs												
General	4.07 (0.73)	585	.68	4.00 (0.69)	588	.70	3.95 (0.72)	572	.77	3.93 (0.71)	574	.73
Math	4.05 (0.89)	590	.83	3.98 (0.88)	586	.86	3.96 (0.96)	572	.90	3.98 (0.88)	574	.87
German	4.14 (0.77)	590	.78	4.04 (0.77)	589	.81	4.02 (0.78)	571	.86	3.97 (0.78)	576	.85

two-way imputation (Enders, 2001; D. A. Newman, 2003).

Recent research demonstrates that the FIML approach is a potent algorithm for handling missing values (Allison, 2001; Collins, Schafer, & Kam, 2001). By means of FIML, rather than missing values being imputed, missing data are recognized as such, and all observed data values are used to estimate models maximizing the power of the analyses (see Allison, 2001).

Latent growth modeling. To test whether intrinsic values and competence beliefs decreased over time (see Research Question 1), latent growth modeling (LGM) procedures were run using Amos 7.0 (McArdle & Epstein, 1987; Stoolmiller, 1995). We chose LGM for several reasons. First, LGM has been found to have more power in detecting changes over time than repeated measures analysis of variance (Fan & Fan, 2005). Second, when using LGM missing data can be handled by FIML. Third, by applying LGM, both inter- and intraindividual differences in changes over time are considered (McArdle & Epstein, 1987). Six models (competence beliefs and intrinsic values in three domains each) were set up with four measurement occasions as well as a correlated latent intercept and slope factor.

SEM. Data were analyzed by computing longitudinal SEM with Amos 7.0. For each of the three domains (math, German, school in general), we set up two competing models to describe the empirical data: one cross-lagged model (Figure 1a) and one correlational model (Figure 1b). Spanning 1 year, the models employed four measurement occasions. The two latent constructs intrinsic values and competence beliefs were indicated by three items each at each measurement occasion.

By means of SEM, the four remaining research questions were addressed:

1. How stable are intrinsic values and competence beliefs over time? This is indicated by the magnitude of the association each constructs shows with itself from one measurement occasion to the next.
2. How strong are the concurrent correlations between the two concepts? This is indicated by the strength of association between intrinsic values and competence beliefs at each measurement occasion.
3. Are prior competence beliefs and subsequent intrinsic values related over time when controlling for prior intrinsic values? And, vice versa, are prior intrinsic values and subsequent competence beliefs related when controlling for prior competence beliefs? The strength of association between the two constructs over time is indicated by the cross-path from one measurement occasion to another. Besides testing the cross-paths running from one measurement occasion to the next, we also tested for cumulative effects from Time 1 to Time 4, that is, over the timespan of 1 year (this was done by three cross-lagged models spanning cross-lagged paths from Time 1 – Time 4).
4. Does a more complex model including cross-lagged paths (Figure 1a) describe the empirical data better than a more parsimonious model with correlational paths but without cross-paths (Figure 1b)? This comparison of the two competing models is a second test of the assumption

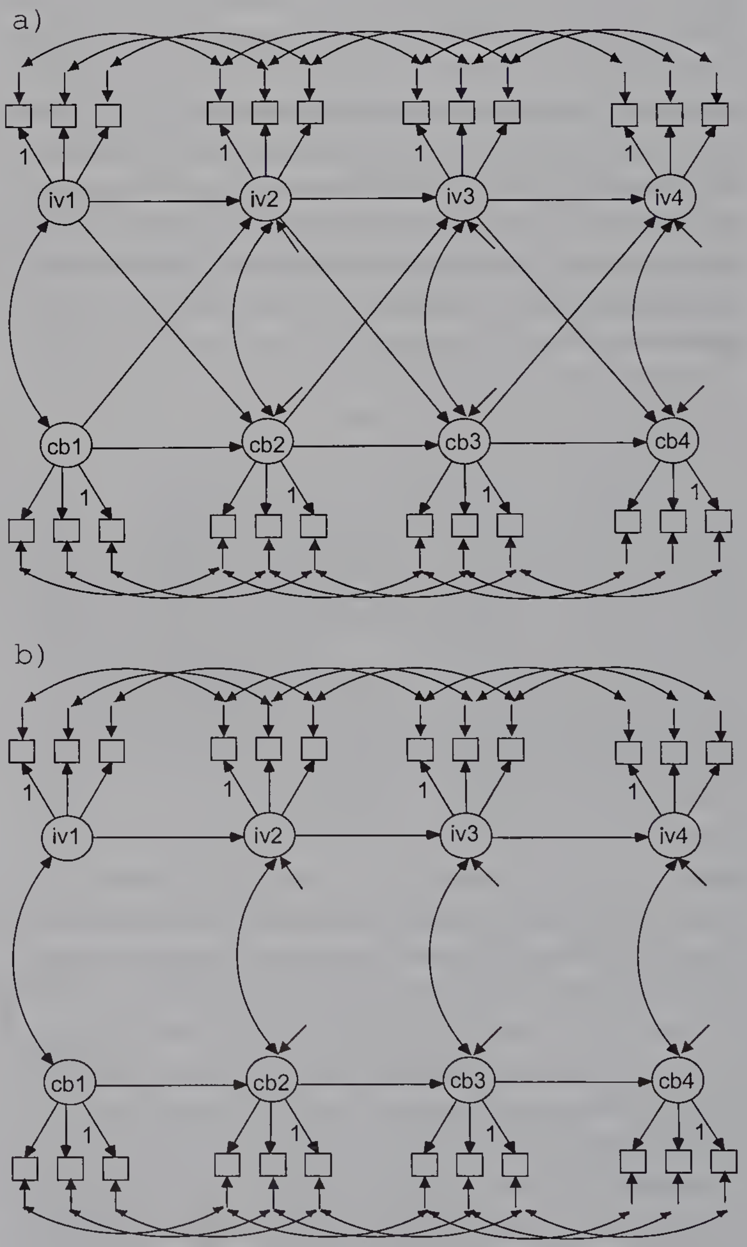


Figure 1. (a) Cross-lagged model: Longitudinal structural equation modeling of intrinsic values (iv) and competence beliefs (cb) including cross-lagged directional paths. (b) Correlational model: Longitudinal structural equation modeling of intrinsic values (iv) and competence beliefs (cb) without cross-lagged directional paths.

> that intrinsic values and competence beliefs might share reciprocal effects over time. According to a basic scientific principle, in the case of two competing models, which describe a phenomenon equally well, the more parsimonious model or theory should be preferred over the more complex one. Although the preference of the more parsimonious model in the case of an equal fit to the data does not rule out the existence of cross-lagged effects, such effects would be rather weak.

Figure 1a depicts the hypothesized cross-lagged effects model (cross-lagged model) employing four measurement occasions. This model allowed for testing further models nested in it. The correlational model (Figure 1b) was modeled with additionally constraining the cross-lagged paths to zero. Thus, the correlational model omits cross-lagged directional paths between the two constructs. The more parsimonious model should be preferred over the more complex model provided that omitting the cross-paths does not deteriorate the model fit (Neale & Cardon, 1992). Differences in fit between nested models were tested by means of likelihood ratio tests. Furthermore, the Tucker – Lewis index was used, a CFI showing the proportional improvement in model fit of one model in comparison to another.

Beside the models employing all four measurement occasions (t1-2-3-4 models), we also tested models for the longest possible time interval, incorporating direct paths from the first to the last measurement occasion (t1-4 models). Here, we also compared t1-4 models with and without cross-lagged effects and named them t1-4 cross-lagged and t1-4 correlational models.

To control for potential memory effects and, thus, positively overestimated stabilities between the different measurement occasions, models were set up with correlated uniqueness between all corresponding measures collected at subsequent measurement occasions. To identify the model, residual variances were constrained.

For the evaluation of overall model fit, three different fit indices were used (see Hu & Bentler, 1999): chi-square value, RMSEA, and CFI. According to Browne and Cudeck (1993), an RMSEA $\leq .05$ indicates a very good model fit and an RMSEA $\leq .09$ is still an indicator for a reasonable error of approximation. According to Hu and Bentler (1995), it is difficult to provide a recommended range for the CFI because in some cases even a CFI < 0.90 can indicate a reasonable model fit. Usually one looks for a CFI ≥ 0.95.

Results

Descriptives

Means, standard deviations, and reliabilities are shown in Table 1. Means of all scales are above the theoretical mean, which signals a potential violation of the normal distribution assumption. In fact, all scales are skewed to the positive pole, as is always the case with these concepts in young children. However, this should not affect the present analyses because maximum likelihood estimations are robust against such violations in the case of large samples, that is, $N > 400$ (see Schermelleh-Engel, Moosbrugger, & Müller, 2003).

Latent Growth Modeling

LGMs were employed to test whether intrinsic values and competence beliefs decreased over time. Results for the six LGMs are depicted in Table 2. The second to fourth columns depict different fit indices for the models (chi-square, RMSEA, and CFI). Model fit for all models was at least satisfactory. The fifth and the sixth columns depict the estimations of the intercept factor and its standard deviations (*M* IC and *SD* IC). The intercept factor represents the average starting point of the construct at t1 and conveys the same information as the means and standard deviations of the first measurement occasion presented in Table 1.

Columns 7 and 8 depict the estimation of the means of the slope factor as well as its standard deviations (*M* Slope and *SD* Slope). The slope factor represents the average change of the constructs over the four measurement occasions. The mean of the slope factor (*M* Slope) was significant for all models with the exception of intrinsic values for German. For these five models, the decline over time was small with scores decreasing by .09 – .15 units. This finding is in line with our assumption that intrinsic values and competence beliefs decline during the early school years.

The standard deviations of the slope factor (*SD* Slope) were significant as well, again with the exception of intrinsic values ascribed to German. This finding indicates interindividual differences in the trajectories.

The ninth column of Table 2 depicts the covariance of the intercept and the slope factor. In accordance with Bast and Reitsma (1997), we do not interpret these covariances because "interpretation of the correlation between status and growth is in most cases not valid" (p. 165) as was also demonstrated by Rovine and Molenaar (1998) by means of simulated data and algebraically.

Table 2

Latent Growth Models for Intrinsic Values and Competence Beliefs in Three Domains

	χ^2	RMSEA	CFI	*M* IC	*SD* IC	*M* Slope	*SD* Slope	Slope × IC
Intrinsic values								
General	4.40	.03	1.00	4.37**	.63**	−.15**	.73**	−.65**
Math	10.00*	.06	0.99	3.86**	.82**	−.09*	.55**	−.24
German	8.56*	.05	0.99	4.09**	.61**	.01	.10	−.18
Competence beliefs								
General	10.39**	.06	0.99	4.05**	.63**	−.16**	.51**	−.51**
Math	23.75**	.10	0.98	4.04**	.77**	−.09*	.47**	−.25*
German	19.00**	.08	0.98	4.13**	.62**	−.15**	.51**	−.38*

Notes. All models listed above are associated with chi-square values with three degrees of freedom. RMSEA = root mean square error of approximation; CFI = comparative fit index; IC = intercept; Slope × IC = covariance between the Slope and Intercept factors.
$^*p < .05$. $^{**}p < .01$.

SEM

Before further interpreting the findings from the SEMs, the general fit of the models to the empirical data needs to be evaluated. Fit indices for the cross-lagged and correlational models with all four measurement occasions (t1-2-3-4) and t1-4 models are depicted in Table 3. CFI and RMSEA fit indices for all models were at least satisfactory, indicating an acceptable to good fit to the data.

Next, we turn to the question of how stable intrinsic values and competence beliefs are over time. Table 4 presents the temporal stabilities for 3-month intervals and for the timespan of 1 year (t1-4). Coefficients are based on maximum likelihood estimations. Considering time intervals of 3 months, temporal stabilities for both constructs were high (ranging from .63 to .87). Over the timespan of 1 year, stabilities were mostly high (ranging from .46 to .70), with the exception of competence beliefs for German (.27). The knowledge that both intrinsic values and competence beliefs are highly stable over time is important because it renders reciprocal effects between the two constructs less probable.

Concurrent correlations between intrinsic values and competence beliefs are of interest because the usually observed strong associations between the two constructs nourish the assumption of mutual influences. Concurrent correlations between intrinsic values and competency beliefs were highest for math (.73 – .83), somewhat lower for German (.60 – .73), and lowest for school in general (.29 – .41), showing that the two constructs are higher correlated within specific domains than for school in general (see Table 5).

Next, we turn to the cross-lagged effects that indicate whether there are reciprocal influences between intrinsic values and competence beliefs. A look at Table 5 reveals that only few significant time lagged cross-paths emerged. Among the cross-paths from competence beliefs to intrinsic values, two of the nine paths reached significance. These two paths were spread over two of the three domains (math and school in general) and two of the three intervals between measurements. Among the cross-paths in the opposite direction, that is from intrinsic values to competence beliefs, none reached significance. When inspecting the longest possible interval between measurements, one of the six cross-paths reached significance, namely, the one from intrinsic values

Table 3

Fit Indices for the Cross-Lagged and Correlational Models Incorporating All Measurement Occasions (t1-2-3-4) and Measurement Occasions 1 and 4 (t1-4) per Domain

	Fit indices				
	χ^2	*df*	χ^2/df	RMSEA	CFI
General					
t1-2-3-4 cross-lagged model	588.92**	223	2.64	.05	0.94
t1-2-3-4 correlational model	598.51**	229	2.61	.05	0.94
t1-4 cross-lagged model	88.60**	42	2.11	.04	0.98
t1-4 correlational model	89.57**	44	2.03	.04	0.98
Math					
t1-2-3-4 cross-lagged model	522.83**	223	2.35	.05	0.96
t1-2-3-4 correlational model	543.25**	229	2.37	.05	0.96
t1-4 cross-lagged model	77.02**	42	1.83	.04	0.99
t1-4 correlational model	85.84**	44	1.95	.04	0.99
German					
t1-2-3-4 cross-lagged model	680.06**	223	3.05	.05	0.92
t1-2-3-4 correlational model	689.13**	229	3.01	.05	0.92
t1-4 cross-lagged model	85.66**	42	2.04	.04	0.97
t1-4 correlational model	93.91**	44	2.13	.04	0.97

Note. RMSEA = root mean square error of approximation; CFI = comparative fit index.
$^{**}p < .01$.

Table 4
Temporal Stabilities (SEM Maximum Likelihood Estimates)

	Temporal stabilities			
	t1 – t2	t2 – t3	t3 – t4	t1 – t4
Intrinsic values				
General	.77**	.73**	.82**	.52**
Math	.66**	.85**	.85**	.70**
German	.71**	.84**	.63**	.60**
Competence beliefs				
General	.74**	.76**	.71**	.63**
Math	.75**	.75**	.87**	.46**
German	.70**	.75**	.70**	.27**

Note. t = measurement occasion; SEM = structural equation modeling.
$^{**}p < .01$.

(t1) to competence beliefs (t4) for German. The magnitude of all significant cross-paths was low. Taken together, these results suggest that there were only minor and unsystematic reciprocal influences between intrinsic values and competence beliefs.

Finally, the correlational models were tested against the cross-lagged models to show whether the more complex cross-lagged models provide a better fit to the data than the mere correlational models. First, we describe the results for the models including all four measurement occasions (t1-2-3-4 models). Concerning school in general and German, the correlational models yielded fit statistics that were as good as those of the cross-lagged models (see Table 6). Omitting the cross-lagged paths did not lead to a deterioration of fit in these two domains. Only in math, the correlational model yielded a deteriorated model fit compared to the cross-lagged model. This effect can be attributed to one single significant path from competence beliefs to intrinsic values from t1 to t2: Constraining this path weight to zero yielded a significant deterioration in fit (discrepancy between chi-squares of cross-lagged and correlational model [CMIN] = 15.93; $df = 1$; $p < .001$), which was not the case when constraining other paths to zero. Therefore, for math, a unidirectional model should be preferred over a correlational, modeling one path from competence beliefs at t1 to intrinsic values at t2. For German and school in general, the cross-lagged paths did not contribute substantially to the model fit. Therefore, the more parsimonious correlational models should be preferred for these two domains.

Turning to the models spanning 1 year (t1-4 models), for school in general, the correlational model yielded no deterioration in fit compared to the cross-lagged model. For German, a deterioration in fit occurred when constraining the path weight from intrinsic values to competence beliefs to zero (CMIN = 7.01; $df = 1$; $p = .01$). In contrast, when constraining the path from competence beliefs to intrinsic values to zero, no significant deterioration in fit occurred (CMIN = .29; $df = 1$; $p = .59$). For math, the model fit deteriorated only when constraining both path weights to zero. Thus, when considering the longest possible timespan of 1 year, for school in general, the correlational model was the most adequate description of the data, whereas for math, a reciprocal cross-lagged model, and for German, a unidirectional model with a cross-path running from intrinsic values to competence beliefs should be preferred.

Taken together, the comparison of model fits of correlational and cross-lagged models suggests that in most cases a correlational model is adequate to describe the empirical data.

Discussion

The main aim of the present study was to shed further light on the question whether the development of intrinsic motivation in the early school years is related to the development of competence beliefs over time. Our study is the first in this field that, simultaneously, incorporates more than two longitudinal

Table 5
Estimated Coefficients (Maximum Likelihood) From Cross-Lagged Domain-Specific SEM Models: Concurrent Correlations and Cross-Path Coefficients

	Concurrent correlations IV × CB				Cross-paths IV × CB							
	t1	t2	t3	t4	IV1 – CB2	IV2 – CB3	IV3 – CB4	IV1 – CB4	CB1 – IV2	CB2 – IV3	CB3 – IV4	CB1 – IV4
General	.41**	.29**	.33**	.36**	.02	−.02	.09	−.06	.01	.10*	−.07	.03
Math	.76**	.83**	.77**	.73**	.06	.04	−.06	.18	.21**	−.01	.00	.19
German	.60**	.73**	.61**	.69**	.07	.03	.01	.26*	.07	−.14	.15	.07

Note. t = measurement occasion; IV = intrinsic value; CB = competence beliefs; SEM = structural equation modeling.
$^{*}p < .05$. $^{**}p < .01$.

Table 6
Model Comparison Between Cross-Lagged and Correlational Models Incorporating All Measurement Occasions (t1-2-3-4) and Measurement Occasions 1 and 4 (t1-4)

	Model comparison			
	df	CMIN	*p*	TLI
General				
t1-2-3-4 models	6	9.59	.14	–.00
t1-4 models	2	.97	.61	–.00
Math				
t1-2-3-4 models	6	20.42	.00	.00
t1-4 models	2	8.82	.01	.00
German				
t1-2-3-4 models	6	9.07	.17	–.00
t1-4 models	2	8.23	.02	.00

Note. CMIN = chi-square discrepancy (between cross-lagged and correlational model); TLI = Tucker–Lewis index.

measurement occasions, addresses different specific domains as well as school in general, and models the constructs on a latent basis. Moreover, we investigated children at a phase in their trajectories, which makes normative evaluations very salient and thus would render the finding of reciprocal relations probable.

Declining Competence Beliefs and Intrinsic Motivation

Consistent with previous findings, our results showed that intrinsic motivation and competence beliefs declined throughout the investigated time period (Bouffard et al., 2003; Gottfried et al., 2001; R. S. Newman, 1984; Spinath & Spinath, 2005b; Wigfield et al., 1997). Additionally, we demonstrated significant interindividual differences in the decline; that is, the decrease in competence beliefs and intrinsic motivation is more pronounced for some children and less for others. This is an important finding because some children seem to be more at risk for losing their motivation to learn. Further studies should focus on variables that might explain these interindividual differences in the developmental curves of intrinsic motivation and competence beliefs. Such factors might have the potential to buffer the decline in intrinsic motivation over time.

Temporal Stability of Competence Beliefs and Intrinsic Motivation

Both intrinsic values and competence beliefs were highly stable in terms of interindividual stability over time. The high stability of the two constructs has direct implications for the search of reciprocal effects because the more stable a construct, the lower the possibility of finding other variables that potentially explain variance (see Eccles, 2005). One might argue that in order to find relations between two constructs over time, it is necessary to investigate the phenomenon as early as possible so that maybe the constructs have not yet reached such high interindividual stabilities. Unfortunately, this approach is limited because competence beliefs in the sense operationalized in our study do not emerge before school age (cf. Dweck, 2002) and seem to be highly stable as soon as they are measurable (e.g., Spinath & Spinath, 2005b). Given the high interindividual stabilities, there is a small chance of finding reciprocal influences between competence beliefs and intrinsic values. Whether this means that there actually are no or only weak reciprocal influences needs to be investigated with different methodological approaches. It is a challenging task for future research to find such methodological approaches.

Concurrent Correlations Between Competence Beliefs and Intrinsic Motivation

Concurrent correlations between competence beliefs and intrinsic values were found to be high for specific domains (math and German) and markedly lower for school in general. Such differences in strength of concurrent associations between constructs are important because one could argue that the probability of finding reciprocal effects is higher in domains in which there are stronger concurrent associations. Stronger associations between competence beliefs and intrinsic motivation within specific domains might be explained by considering that when children think about competence and task values in a certain subject, they think of the same material and teacher. However, regarding school in general, one topic may be more strongly reflected when children think about their intrinsic values (e.g., art class) and another when they think of their competence (e.g., geography). This might be a possible explanation for the lower concurrent correlations in the general domain. This result indicates that it is worthwhile to include different domains in one study because the probability of finding reciprocal effects might depend on the degree of association between the constructs.

The reason why intrinsic motivation and competence beliefs are found to be so highly correlated, although they seem not to be linked over time, might lie in the fact that certain experiences, like receiving unfavorable feedback, often affect both competence beliefs and intrinsic values at the same time. Nevertheless, in such situations, decreasing competence beliefs could be independent from decreasing intrinsic values. Even at instances when competence beliefs might realistically decrease after negative feedback, children's intrinsic motivation can stay at the same level or even increase when they are oriented toward learning and when they are convinced that effort will help to overcome initial difficulties. Thus, interventions that focus children on individual progress, learning goals and strengthen their faith in the utility of effort might help to disentangle the developmental curves of competence beliefs and intrinsic values.

Relations Between Competence Beliefs and Intrinsic Motivation Over Time

Concerning the question of potential relations between intrinsic values and competence beliefs over time, our results are in line with previous studies showing that such effects are, if observable at all, of negligible magnitude (Marsh et al., 2005; Nurmi & Aunola, 2005; Skaalvik & Valas, 1999; Spinath & Spinath, 2005b). We conclude this from the observations that, in most cases, mere correlational models provided as good descriptions of the data as more complex cross-lagged models and that cross-paths between intrinsic values and competence beliefs, overall, were very weak. The few significant relations between constructs over time revealed no interpretable pattern: Of 24 cross-paths, only 3 reached significance, which were spread over all three domains as well as different measurements. One path did not run from prior competence beliefs to subsequent intrinsic values but the other way around. All in all, our results indicate that the probability of a relation between ability beliefs and intrinsic values over time is low. After controlling for prior intrinsic values, there are only negligible associations between competence beliefs and subsequent intrinsic values and vice versa.

Before this conclusion is accepted, one needs to exclude methodological reasons, which might have prevented finding reciprocal effects. We took great efforts to design the present study in line with methodological and content-specific recommendations made by experts in the field (Eccles, 2005; Marsh & Yeung, 1997, 1998). Such efforts included longitudinal measurement with four measurement occasions, short intervals between measurements, a midsized sample of elementary school children, latent measurement of constructs, and incorporating different domains. Especially noteworthy is that we investigated children at a period during their school trajectories where normative evaluations become salient because of the onset of grades. Taken together, we believe that those features of the present study rendered the finding of reciprocal relations over time between competence beliefs and intrinsic values especially probable so that neither methodological nor other characteristics of the study account for not finding such associations over time.

Implications for Motivation Theories and Educational Practice

Given the fact that many motivation theories implicitly or explicitly assume that positive competence beliefs are a very important prerequisite for intrinsic motivation and interest (Deci & Ryan, 1985; Harter, 1981a; Wigfield & Eccles, 2000), our results are surprising. We agree that it is a very plausible assumption that individuals should enjoy tasks, which give them reasons to feel competent. But we like to advance the idea that it is not the absolute level of normatively based ability self-perceptions that stimulates feelings of efficacy but other ability-related perceptions and beliefs. Specifically, we argue that perceptions of individual progress and the optimism that competences can be developed should be important to sustain intrinsic motivation in a certain domain.

These kinds of ability-related beliefs have been discussed in different developmentally focused approaches to motivation. For example, Nicholls (1984) distinguishes between two ways in which to conceive ability: Ability can be judged relative to others or with reference to the individual's own past performance. We suggest that children who measure their own competence more often with reference to their past performance instead of comparing it with others' ability should be more strongly intrinsically motivated in the corresponding tasks. A second ability-related concept we want to suggest as a promising candidate for fostering intrinsic motivation are implicit theories about the malleability of ability. The work of Dweck and her colleagues (cf. Dweck, 1999) shows that implicit theories about the malleability of ability are largely independent of the absolute level of self-perceived ability. We expect that children who are convinced that their ability is not a fixed entity but can be increased should be more intrinsically motivated to learn and thus actually increase their competence. To corroborate these theoretical assumptions, future studies should include measures such as perceptions of individual progress and incremental ability theories when investigating influences on intrinsic motivation (see also Spinath & Spinath, 2005b).

With respect to educational practice, the present results allow some speculations about more and less effective methods for fostering high, sustaining intrinsic motivation. Many teachers hold the assumption that declining intrinsic motivation to learn is an inevitable consequence of children's increasingly realistic self-perceptions. Some teachers might therefore try and preserve children's unrealistic ability self-perception by giving either unrealistically positive or vague feedback. Given the present findings, such practices will probably fail to conserve children's intrinsic motivation. The development of realistic ability self-perceptions is an important developmental task, which needs to be supported by realistic feedback. If the decline of intrinsic motivation in school-related learning is not linked to declining competence beliefs, then it should be possible to disentangle the developmental curves of the two constructs. Even in the face of realistically held low ability self-concepts, learners can develop an optimistic, learning-oriented perspective in which they consider low competencies as learning opportunities and learning as an end in itself.

References

Allison, P. D. (2001). *Missing data*. Thousand Oaks, CA: Sage.

Bast, J., & Reitsma, P. (1997). Matthew effects in reading: A comparison of latent growth curve models and simplex models with structured means. *Multivariate Behavioral Research*, *32*, 135–167.

Bouffard, T., Marcoux, M. F., Vezeau, C., & Bordeleau, L. (2003). Changes in self-perceptions of competence and intrinsic motivation among elementary school children. *British Journal of Educational Psychology*, *73*, 171–186.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Beverly Hills, CA: Sage.

Cocodia, E. A., Kim, J.-S., Shin, H.-S., Kim, J.-W., Ee, J., Wee, M. S. W., et al. (2003). Evidence that rising population intelligence is impacting in formal education. *Personality and Individual Differences*, *35*, 797–810.

Collins, L. M., Schafer, J. L., & Kam, C.-M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods*, *6*, 330–351.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum.

Dweck, C. S. (1999). *Self-theories: Their role in motivation, personality, and development*. Philadelphia: Taylor & Francis.

Dweck, C. S. (2002). The development of ability conceptions. In A. Wigfield & J. S. Eccles (Eds.), *Development of achievement motivation* (pp. 57–88). San Diego, CA: Academic Press.

Eccles, J. S. (2005). Studying the development of learning and task motivation. *Learning and Instruction*, *15*, 161–171.

Eccles, J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., & Meece, J. L., et al. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motives* (pp. 75–146). San Francisco: Freeman.

Eccles, J. S., Wigfield, A., Harold, R. D., & Blumenfeld, P. (1993). Age and gender differences in children's self- and task perceptions during elementary school. *Child Development*, *64*, 830–847.

Enders, C. K. (2001). The impact of nonnormality on full information maximum-likelihood estimation for structural equation models with missing data. *Psychological Methods*, *6*, 352–370.

Fan, X., & Fan, X. (2005). Power of latent growth modelling for detecting linear growth: Number of measurements and comparison with other analytic approaches. *Journal of Experimental Education*, *73*, 121–139.

Gottfried, A. E. (1990). Academic intrinsic motivation in young elementary school children. *Journal of Educational Psychology*, *82*, 525–538.

Gottfried, A. E., Fleming, J. S., & Gottfried, A. W. (2001). Continuity of academic intrinsic motivation from childhood through late adolescence: A longitudinal study. *Journal of Educational Psychology*, *93*, 3–13.

Harter, S. (1981a). A model of mastery motivation in children: Individual differences and developmental change. In W. A. Collins (Ed.), *Aspects on the development of competence: The Minnesota Symposia on Child Psychology* (Vol. 14, pp. 215–255). Hillsdale, NJ: Erlbaum.

Harter, S. (1981b). A new self-report scale of intrinsic versus extrinsic orientation in the classroom: Motivation and informational components. *Developmental Psychology*, *17*, 300–312.

Harter, S. (1982). The perceived competence scale for children. *Child Development*, *53*, 87–97.

Howard, R. W. (2001). Searching the real world for signs of rising population intelligence. *Personality and Individual Differences*, *30*, 1039–1058.

Hu, L.-T., & Bentler, P. M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modelling: Concepts, issues, and applications* (pp. 76–99). London: Sage.

Hu, L.-T., & Bentler, P. M. (1999). Cut-off criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55.

Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. *Child Development*, *73*, 509–527.

Marsh, H. W. (1990). Causal ordering of academic self-concept and academic achievement: A multi-wave, longitudinal panel analysis. *Journal of Educational Psychology*, *82*, 646–656.

Marsh, H. W., Byrne, B. M., & Yeung, A. S. (1999). Causal ordering of academic self-concept and achievement:

Reanalysis of a pioneering study and revised recommendations. *Educational Psychologist, 34*, 155–167.

Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2005). Academic self-concept, interest, grades, and standardized test scores: Reciprocal effects models of causal ordering. *Child Development, 76*, 397–416.

Marsh, H. W., & Yeung, A. S. (1997). Causal effects of academic self-concept on academic achievement: Structural equation models of longitudinal data. *Journal of Educational Psychology, 89*, 41–54.

Marsh, H. W., & Yeung, A. S. (1998). Top-down, bottom-up, and horizontal models: The direction of causality in multidimensional, hierarchical self-concept models. *Journal of Personality and Social Psychology, 75*, 509–527.

McArdle, J. J., & Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development, 58*, 110–133.

Neale, M. C., & Cardon, L. R. (1992). *Methodology for genetic studies of twins and families*. Boston: Kluwer Academic.

Newman, D. A. (2003). Longitudinal modelling with randomly and systematically missing data: A simulation of ad hoc, maximum likelihood, and multiple imputation techniques. *Organizational Research Methods, 6*, 328–362.

Newman, R. S. (1984). Children's achievement and self-evaluations in mathematics: A longitudinal study. *Journal of Educational Psychology, 76*, 857–873.

Nicholls, J. G. (1978). The development of the concepts of effort and ability, perception of academic attainment, and the understanding that difficult tasks require more ability. *Child Development, 49*, 800–814.

Nicholls, J. G. (1984). Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. *Psychological Review, 91*, 328–346.

Nicholls, J. G., & Miller, A. T. (1984). Reasoning about the ability of self and others: A developmental study. *Child Development, 55*, 1990–1999.

Nurmi, J.-E., & Aunola, K. (2005). Task-motivation during the first school years: A person-oriented approach to longitudinal data. *Learning and Instruction, 15*, 103–122.

Rovine, M. J., & Molenaar, P. C. M. (1998). The covariance between level and shape in the latent growth curve model with estimated basis vector coefficients. *Methods of Psychological Research Online, 3*, 95–107.

Ruble, D. N., Boggiano, A. K., Feldman, N. S., & Loebl, J. H. (1980). Developmental analysis of the role of social comparison in self-evaluation. *Developmental Psychology, 16*, 105–115.

Ruble, D. N., Parsons, J. E., & Ross, J. (1976). Self-evaluative responses of children in an achievement setting. *Child Development, 47*, 990–997.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online, 8*, 23–74.

Skaalvik, E. M., & Valas, H. (1999). Relations among achievement, self-concept, and motivation in mathematics and language arts: A longitudinal study. *Journal of Experimental Education, 67*, 135–149.

Spinath, B., & Spinath, F. M. (2005a). Development of self-perceived ability in elementary school: The role of parents' perceptions, teacher evaluations, and intelligence. *Cognitive Development, 20*, 190–204.

Spinath, B., & Spinath, F. M. (2005b). Longitudinal analysis of the link between learning motivation and competence beliefs among elementary school children. *Learning and Instruction, 15*, 87–102.

Stipek, D. J., & Daniels, D. H. (1988). Declining perceptions of competence: A consequence of changes in the child or in the educational environment? *Journal of Educational Psychology, 80*, 352–356.

Stipek, D. J., & MacIver, D. (1989). Developmental change in children's assessment of intellectual competence. *Child Development, 60*, 521–538.

Stoolmiller, M. (1995). Using latent growth curve models to study developmental processes. In J. M. Gottman (Ed.), *Analysis of developmental change* (pp. 103–138). Hillsdale, NJ: Erlbaum.

White, R. W. (1959). Motivation reconsidered: The concept of competence. *Psychological Review, 66*, 297–333.

Wigfield, A. (1994). Expectancy-value theory of achievement motivation: A developmental perspective. *Educational Psychology Review, 6*, 49–78.

Wigfield, A., & Eccles, J. S. (1992). The development of achievement task values: A theoretical analysis. *Developmental Review, 12*, 265–310.

Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology, 25*, 68–81.

Wigfield, A., Eccles, J. S., Yoon, K. S., Harold, R. D., Arbreton, A. J. A., & Freedman-Doan, C., et al. (1997). Change in children's competence beliefs and subjective task values across the elementary school years: A 3-year study. *Journal of Educational Psychology, 89*, 451–469.

Child Development, September/October 2008, Volume 79, Number 5, Pages 1570–1582

Children's Insecure Representations of the Interparental Relationship and Their School Adjustment: The Mediating Role of Attention Difficulties

Patrick T. Davies and Meredith J. Woitach
University of Rochester

Marcia A. Winter
Syracuse University

E. Mark Cummings
University of Notre Dame

This study examined the role of attention difficulties as a mediator of associations between children's insecure representations of the interparental relationship and their school adjustment in a sample of two hundred and sixteen 6-year-old children. Consistent with hypotheses, findings from structural equation models indicated that observer ratings of children's insecure representations of interparental relationships in a story completion task predicted computerized task assessments and parent reports of children's attention difficulties 1 year later. Children's attention difficulties, in turn, were associated with concurrent levels of school problems and increases in school problems over a 1-year period as indexed by teacher reports. Attention difficulties accounted for an average of 34% of the association between insecure internal representations and school problems.

The significance of interparental discord as a public health concern is highlighted by its association with children's psychological maladjustment. Exposure to high levels of discord increases children's risk for a wide array of psychological problems including internalizing symptoms (e.g., depression, anxiety), externalizing problems (e.g., aggression), social impairments (e.g., poor peer relations), and academic difficulties (e.g., Dunn & Davies, 2001; Erel & Burman, 1995; Grych & Fincham, 2001). The disproportionate vulnerability of children exposed to interparental discord has precipitated calls to understand precisely how and why interparental conflict is associated with children's psychological functioning.

In addressing why interparental strife is a risk factor for children, prevailing theories propose that children's internal representations of the consequences of interparental difficulties are pivotal explanatory mechanisms. According to the emotional security theory (EST; Cummings & Davies, 1996), witnessing destructive interparental conflict sensitizes children to concerns about preserving their security. Threats to security are manifested in children's negative internal representations of the consequences for the welfare of the self and family. Although these representations may be adaptive initially by sharpening children's ability to identify danger cues in high-conflict homes, maladaptive implications for their long-term psychological adjustment are expected (Davies, Winter, & Cicchetti, 2006; Johnston & Roseby, 1997). Similarly, the cognitive-contextual framework hypothesizes that elevated interparental conflict increases children's vulnerability to psychological problems by heightening their appraisals of threat in the family (Grych & Fincham, 1990). Supporting the common prediction of these theories, cross-sectional and longitudinal studies indicate that insecure representations of interparental relations are intermediary mechanisms in associations between interparental conflict and child adjustment (Dadds, Atkinson, Turner, Blums, & Lendich, 1999; Davies & Cummings, 1998; Grych, Fincham, Jouriles, & McDonald, 2000; Grych, Harold,

---

This research was supported by the National Institute of Mental Health (Project R01 MH57318) awarded to Patrick T. Davies and E. Mark Cummings. We are grateful to the children, parents, teachers, and school administrators who participated in this project. Our appreciation is expressed to project staff, including Courtney Forbes, Courtney Henry, Marcie Goeke-Morey, Amy Keller, Michelle Sutton, Alice Schermerhorn, and the graduate and undergraduate students at the University of Rochester and University of Notre Dame. We would also like to thank Melissa Sturge-Apple for her statistical consultation and expertise.

Correspondence concerning this article should be addressed to Patrick T. Davies, Department of Clinical and Social Sciences in Psychology, University of Rochester, Rochester, NY 14627. Electronic mail may be sent to patrick.davies@rochester.edu.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7905-0023

& Miles, 2003; Harold, Shelton, Goeke-Morey, & Cummings, 2004).

Although this work documents the importance of negative representations for understanding the development of children exposed to interparental discord, questions of how and why these representations develop into broader patterns of psychological maladjustment remain unexplored. To address this gap, this study examines attention difficulties as possible mediating mechanisms in accounting for links between children's negative representations of interparental relations and their psychological adjustment during the early elementary school years. Guided by a common theoretical focus on the evaluation of threat and safety in internal representation constructs (Davies & Cummings, 1994; Grych & Fincham, 1990), we trace the sequelae of children's insecure representations of interparental relations or, more specifically, their analysis of the threatening repercussions of interparental difficulties for their well-being.

In addressing the first link in our proposed mediational chain, recent conceptual advances in EST provide the groundwork for outlining how children's insecure representations of interparental relations may increase children's risk for experiencing attention difficulties (Davies, Winter, et al., 2006). EST proposes that children's insecure representations of interparental relationships reflect their heightened concerns about preserving their safety and security in the family. Prolonged concerns about security, in turn, set in motion multiple defensive processes that are hypothesized to increase children's risk for adjustment problems by depleting children of biopsychological resources. As a result, insecure children may experience problems with self-regulation across multiple domains, including emotion, cognition, and motivation. Given that directing and sustaining attention require careful coordination of cognitive, affective, regulatory, and motivational processes, concerns about security undermine children's attention. Therefore, insecure representations of interparental relationships may increase children's adjustment problems by disrupting their ability to focus and sustain attention.

Associations between negative family experiences and attention difficulties have been documented. For example, Fearon and Belsky (2004) reported that children with insecure parent–child relationships evidenced elevated attention difficulties under conditions of high cumulative risk (i.e., child, social, and parent risk factors). Likewise, Carlson, Jacobvitz, and Sroufe (1995) found that maternal perceptions of low emotional support from their intimate partners and social network during infancy predicted preschool children's distractibility 2 years later. Thus, even though individual differences in attention difficulties may have constitutional origins that are relatively stable over time, a growing corpus of research supports the role of family environmental factors as precursors to attention difficulties (Groot, de Sonneville, Stins, & Boomsa, 2004; National Institute of Child Health and Human Development [NICHD] Early Child Care Research Network, 2003, 2005).

In the second link in our mediational model, inattention is regarded as a precursor of children's school adjustment problems. For example, attention difficulties may be manifested in impairments in processing relevant contextual cues that act as guides for skillful and appropriate behavioral conduct in social and academic settings. When children are faced with the challenges of adapting to elementary school, attention impairments may contribute to difficulties in establishing harmonious peer relations, exercising self-control, and assuming a cooperative, engaged role in structured classroom contexts (Barkley, Fischer, Smallish, & Fletcher, 2002; Campbell, Pierce, March, Ewing, & Szumowski, 1994). Within clinical and high-risk samples, children with attention difficulties are at greater risk for problems with behavioral conduct, peer relationships, academic achievement, and social engagement (Barkley et al., 2002; Campbell et al., 1994). In community samples, attention impairments during preschool and early school years are identified as concurrent and prospective predictors of lower academic achievement, poorer social competence, and greater externalizing symptoms (NICHD Early Child Care Research Network, 2003). In addition, preschool and elementary school children with difficulties in focusing and shifting attention are at risk for experiencing externalizing symptoms (Eisenberg et al., 2000), poor adaptation to changing demands (Eisenberg, Fabes, Guthrie, & Reiser, 2000), social skill impairments (Eisenberg et al., 1993), and peer problems (Eisenberg et al., 1993).

Our decision to examine mediational pathways among insecure representations of interparental relations, attention difficulties, and child psychological adjustment during the early elementary school years was guided by several considerations. First, significant advances in social perspective taking during this developmental period may elevate the capacity for child worries about family relationships. Thus, internal representations of interparental relationships may assume greater salience in school children's lives as they become increasingly adept at assessing the implications of interparental discord for the welfare of themselves and the family system (Cicchetti, Cummings, Greenberg, & Marvin, 1990). Second, these changes correspond temporally with the emergence and intensification of significant difficulties in sustained attention (Loeber, Green, Lahey, Christ, & Frick, 1992). Accordingly, the early school years provide an informative window for exploring attention difficulties as possible sequelae of negative internal representations. Finally, during this period, children are also faced with the stage-salient task of adjusting to multiple scholastic challenges including regulating behavior, establishing and maintaining peer relationships, and engaging cooperatively in academic activities (Entwisle, 1995; Ladd, Herald, & Andrews, 2006). Because the demands of schooling pose formidable challenges even for children developing along adaptive trajectories, the liabilities associated with insecure representations and attention difficulties for children may be prominent within domains of school adjustment. Thus, this is a critical developmental period for exploring the mediational interplay among representations and attention difficulties in predicting school adjustment.

Breaking new ground methodologically in the study of process variables underlying the impact of representations, we tested the role of attention problems as a mediator of negative representations of interparental relationships using three measurement occasions, each spaced 1 year apart. Employing three measurement occasions generated rigorous tests of our mediator predictions by temporally disaggregating the putative risk factor, mediator, and outcome (Cole & Maxwell, 2003). Therefore, temporally ordered prospective analyses were possible using Wave 1 measures of insecure interparental representations, Wave 2 assessments of attention difficulties, and Wave 3 assessments of school adjustment. Autoregressive analyses, which controlled for prior values of school adjustment and inattention, were conducted to derive more rigorous tests of change in testing mediational pathways (Gollob & Reichardt, 1987). The multimethod, multiinformant measurement battery reduced common method and informant variance and increased confidence for detecting substantively meaningful relationships among the constructs.

Method

Participants

Participants were drawn from a larger project that originally included 235 families recruited through local school districts and community centers in a moderate-sized metropolitan area in the Northeast and a small city in the Midwest. Interested families were included in the project if they met the following eligibility criteria: (a) the primary caregivers had a child in kindergarten, (b) the kindergarten child and two primary caregivers lived together for at least the preceding 3 years, and (c) the primary caregivers and child were fluent in English. Although the criteria did not exclude families with same-sex caregivers, each family in our sample contained a female caregiver, a male caregiver, and a child in kindergarten. Only families who participated in all three waves of data collection were included in this study. The retention rate across all three measurement occasions was 92%. Thus, the final sample for this article consisted of 216 children and their families. The average age of the children at Wave 1 was 6.0 years ($SD = 0.48$); 54% of the sample were girls ($n = 117$) and 46% were boys ($n = 99$).

Median family income for participating families was in the range of $40,000–$54,000, with 11% of the sample reporting household income below $23,000. The average number of years of education completed by mothers and fathers were 14.62 ($SD = 2.25$) and 14.76 ($SD = 2.71$), respectively. A large proportion of the family members reported being White (77%), followed by smaller percentages of Black (16%), Hispanic (3%), Asian American (1%), mixed (2%), and other racial (1%) family members. Families lived together for at least 3 years prior to initial data collection ($M = 5.35$, $SD = 1.06$). Children lived with their biological mother in the majority of cases (95%), with the remainder of the sample living with either an adoptive mother (3%) or a stepmother or female guardian (2%). In addition, children lived with their biological father in the majority of cases (88%), with the remainder of the sample living with either an adoptive father (4%) or a stepfather or male guardian (8%). Minimal differences were found between the present sample ($n = 216$) and the families in the larger sample who did not participate in all waves of the study ($n = 19$). Statistical comparisons of the two groups along 15 demographic variables (e.g., family income, occupational status, race–ethnicity, marital status, age) and the 12 child measures at Waves 1 and 2 (i.e., indices of child insecurity, inattention, and school adjustment) yielded one significant difference: Families in the present sample contained parents with higher education levels than excluded families (Cohen's $d = 0.86$).

Procedures

Data for this study were collected at three measurement occasions, with contiguous measurement occasions spaced 1 year apart. At each wave, families visited the laboratory at one of the two sites. Laboratories at each site were designed to be comparable to each other in size and quality and included (a) an observation room that was designed to resemble a family room and equipped with audiovisual equipment to capture family interactions and (b) interview rooms for completing confidential interview and survey measures. The study was approved by the institutional review board at each research site.

At Wave 1, parents completed abbreviated questionnaires designed to assess attention difficulties. During the Wave 1 visits, children also participated in a revised version of the MacArthur Story Stem Battery (MSSB), a narrative storytelling technique designed to assess child representations of family relationships (Bretherton, Oppenheim, Buchsbaum, Emde, & the MacArthur Narrative Group, 1990). Consistent with the original format of the MSSB, the experimenter conducting the revised version of the MSSB (MSSB–R) introduced a series of story stems describing various stressors and threats to the different family subsystems (Davies, Sturge-Apple, Winter, Cummings, & Farrell, 2006). The MSSB–R is designed to provide a more extensive measurement of child representations of the family system, including three stories each being devoted to assessing threats to the mother–father, mother–child, and father–child relationships. To facilitate child engagement in the task, experimenters used dramatic, animated voices, various toy props, and family action figures matching the child's sex and race. After the experimenter presented each story stem, children completed the story with the assistance of the action figures, props, and various probes and prompts by the experimenter. Given our focus on the security of children's representations of interparental relations, for this study, we only retained the three stories depicting various levels of threat or stress in the interparental relationship: (a) a mild interparental conflict regarding a lost set of keys, (b) an intense interparental conflict regarding a messy kitchen, and (c) a relatively calm interparental disagreement about one of the parents returning home late. Child stories were videotaped for later coding.

At Wave 2, data on attention difficulties were collected with multiple methods and from multiple informants. First, mothers and fathers independently reported on children's attention difficulties through questionnaires. Second, the Continuous Performance Test II (CPT II) was completed by children at Wave 2 to assess their attention difficulties (Conners, 2000). Administration followed standard practices for the task: With an experimenter in the room to oversee the task, children were instructed to watch a series of letters and respond to all the letters except the target letter X by pressing the space bar as quickly as possible. Each letter was displayed for a 250-ms time period. To obtain a more sensitive measure of attention difficulties in the face of the changing demands of the task, time between stimulus (i.e., letter) presentations in the CPT II varied across the task at 1-, 2-, and 4-s intervals. Following a 2-min practice session, children completed six blocks of letter presentations, each lasting approximately 2 min. Each block, in turn, contained three subblocks of 20 trials each.

Finally, children's classroom teachers completed questionnaires at Waves 2 and 3 to assess children's behavioral and social adjustment in school.

Measures

The measurement battery was designed to generate latent constructs of five primary constructs: children's insecure representations of interparental relations at Wave 1, their attention difficulties at Waves 1 and 2, and their school adjustment at Waves 2 and 3. Therefore, each main construct was assessed through multiple indicators.

Insecure internal representations of interparental relations (Wave 1). Coders rated the videotaped records of each of the three interparental narratives from the MSSB–R along four 5-point scales designed to assess child representations of the implications of interparental difficulties for the welfare of the child and family. *Maternal and paternal caregiver incompetence* ratings, which served as two of the indicators of children's insecure representations of interparental relations, were designed to assess the extent to which each caregiver fails to use available resources in ways that protect or improve the integrity of the family. Whereas *very competent* (1) ratings reflect that the child portrays caregivers as sources of support and protection for the child and family in the face of interparental challenges, *very incompetent* (5) ratings reflect that caregivers are depicted as completely ineffective or frightening. As a third indicator, ratings of *relationship quality* were designed to assess children's representations of the long-term emotional impact of the conflict on the interparental relationship. The rating scale ranged from *intense harmony* (1), which reflects expectations that relationships are predominantly harmonious, to *intense discord* (5), reflecting expectations of long-term intense problems between parents. The final indicator, the *emotional insecurity* code, was a global rating designed to assess children's collective representation of the interparental relationship as a source of support or threat. Ratings ranged from *strong security* (1), in which the family relationship is depicted as competently resolving challenges in a manner that facilitates family harmony and the physical and emotional welfare of the child, to *strong insecurity* (5), in which the family relationship is portrayed as a long-term severe threat to the child's security.

Four independent coders, who were trained to reliability, were randomly assigned to code MSSB–R narratives of subsamples of children in the study. To evaluate interrater reliability, all coders were responsible for rating the same subsample (i.e., 20%) of tapes. Intraclass correlation coefficients for each of the four rating scales within each story ranged from .85 to .95 (M = .90).

Attention difficulties (Wave 1). Although our measurement battery did not contain extensive measures of attention difficulties at Wave 1, we used mother and father reports on two questionnaire items as proxy measures of child attention difficulties (i.e., "has a hard time following instructions"; "is good at following instructions"). The items were adapted from the Inhibitory Control Scale of the Child Behavior Questionnaire (CBQ) based on their conceptual correspondence with attention problems (Rothbart, Ahadi, Hersey, & Fisher, 2001). Response alternatives ranged from *extremely untrue of your child* (1) to *extremely true of your child* (7). After recoding the items so that higher scores reflect poorer attention, the four items were averaged together to create a single composite of attention difficulties. The four-item measure evidenced satisfactory internal consistency (α = .73).

Attention difficulties (Wave 2). Attention difficulties at Wave 2 consisted of four indicators. Two measures were derived from the CPT to obtain a balanced, but parsimonious, assessment of children's attention difficulties. First, the *omission error* score was calculated as the percentage of nonresponses to all letters except the target letter X. Thus, higher scores reflect greater difficulties in identifying and responding to the visual stimuli. Second, *variability of standard error*—an assessment of consistency of response times to target stimuli across the test—was defined as the standard deviation of the standard error values of each of the 18 subblocks. Higher scores on this dimension correspond with greater variability in performance across the task and difficulties with sustained attention. The three measures are indexed by T scores that characterize the performance of the children in relation to the normative group according to gender and age. Each of the CPT variables has been shown to have adequate reliability (i.e., split half, test–retest; Conners, 2000). The validity of the CPT assessment is supported by its ability to predict psychological adjustment difficulties (e.g., Conners, 2000; Reeves et al., 2006; Seidel & Joschko, 1990; Zelkowitz, Paris, Guzder, & Feldman, 2001).

The remaining two indicators of attention difficulties were obtained from mother and father reports on the Attentional Focusing scale from the CBQ (Rothbart et al., 2001). The nine items of the scale assess children's temperamental disposition to direct and sustain attention during tasks and activities (e.g., "When picking up toys or other jobs, usually keeps at the task until it's done"; "When drawing or coloring in a book, shows strong concentration"). Response alternatives ranged from *extremely untrue of your child* (1) to *extremely true of your child* (7). Prior research provides evidence for the internal consistency and interrater reliability of the Attentional Focusing scale (Putnam & Rothbart, 2006; Rothbart et al., 2001). In support of its discriminant validity, psychometric analyses revealed no measurement confounding (i.e., item overlap) between the Attentional Focusing scale and psychological symptoms. Likewise, consistent with the assumption that individual differences in temperament may increase the likelihood of child psychological problems, convergent validity of the scale is documented by its power to consistently predict adjustment problems and psychological symptoms (Lemery, Essex, & Smider, 2002). In this sample, alpha coefficients for mother and father reports on the Attentional Focusing scale were .73 and .70, respectively. To be consistent with the CPT assessments, maternal and paternal reports were reverse-scored so that higher values reflect greater attention difficulties.

Children's school adjustment. At Waves 2 and 3, teacher reports on three questionnaire scales were used as indicators of school adjustment. First, to assess the peer component of school adjustment, teachers completed the Prosocial subscale from the Child Behavior Scale (Ladd & Profilet, 1996). The Prosocial subscale contains seven items designed to assess children's supportive behavior toward peers (e.g., "Cooperative with peers"). Teachers responded to each item by selecting one of the three response alternatives (1 = *doesn't apply*, 2 = *applies sometimes*, and 3 = *certainly applies*). Reliability and validity of the measure are supported by several studies (Clark & Ladd, 2000; Ladd & Profilet, 1996). Internal consistency of the scale in our study was satisfactory (αs = .91 and .87 at Waves 2 and 3, respectively).

Second, teacher reports on the Cooperative Participation subscale of the Teacher Rating Scale of School Adjustment were used to assess children's adjustment to classroom demands (Birch & Ladd, 1997). The subscale consists of the sum of seven items designed to assess the extent to which children comply with classroom rules and standards in a responsible, respectful manner (e.g., "Follows teacher's directions"; "Uses classroom materials responsibly"). Response alternatives included *doesn't apply* (0), *sometimes applies* (1), and *certainly applies* (2). Considerable empirical documentation supports the psychometric properties of the Cooperative Participation subscale (e.g., Birch & Ladd, 1997; Ladd & Burgess, 2001). For this sample, alpha coefficients for the subscale for Waves 2 and 3 were .91 and .87, respectively.

Third, teachers completed the Behavioral Competence subscale from the Teacher's Rating Scale of Actual Child Behavior (TRS; Harter & Pike, 1984). The Behavioral Competence scale consists of three items measuring children's behavioral conduct in school. Each item is organized in a sequence of two alternative statements (e.g., "This child usually acts appropriately" or "This child would be better if s/he acted differently"). After selecting which statement best characterized the child, teachers then rated the degree of correspondence between the child and the statement by selecting either "sort of true" or "really true." Responses are quantified along a 4-point scale where 4 corresponds with the endorsement of "really true" to the behaviorally competent statement and 1 signifies the endorsement of "really true" to the behaviorally incompetent statement. The factor structure, internal consistency, and validity of the Behavioral Competence subscale are documented by prior studies (e.g., Harter & Pike, 1984; see also Boivin & Begin, 1989; Webster-Stratton & Lindsay, 1999). For this study, alpha coefficients for the TRS subscale were .91 at Wave 2 and .93 at Wave 3.

Results

Descriptive Analyses

Table 1 depicts the means, standard deviations, and correlations among the main variables in the primary analyses. The means for teacher assessments of child school adjustment were comparable to other large nonclinical and community samples (e.g., Kochenderfer & Ladd, 1996; Ladd & Profilet, 1996; Webster-Stratton & Lindsay, 1999). Likewise, in our sample, mean *T* scores in the low to middle 50s for the three dimensions of attention on the CPT fell within the average (i.e., 45–54) to mildly atypical (i.e., 55–59) range. Correlations among the proposed manifest indicators of the main constructs in Table 1 provided initial support for the operationalization of our measurement model. Correlations among the indicators of each of the constructs were generally moderate to strong in magnitude (see bolded coefficients in Table 1). In support of our proposed pathways, significant interrelationships were evident among Wave 1 dimensions of children's insecure representations, their Wave 2 attention difficulties, and teacher reports of their school adjustment at Waves 2 and 3.

Primary Analyses

Given the significant correlations among observed measures of insecure internal representations of interparental relations, attention difficulties, and school adjustment, we proceeded to test the role of attention difficulties as an explanatory mechanism using structural equation modeling (SEM). SEM analyses were conducted using Amos 7.0 (Arbuckle, 2006), a computer software package that uses the full information maximum likelihood approach to handle missing data. Manifest indicators corresponding to teacher reports on the same subscales were permitted to correlate across measurement occasions for all models specified. To obtain a multimethod, multiinformant index of attention difficulties, the three computer assessments and two parent report measures of attention problems were specified as indicators of a single latent construct. In support of fit of the measurement model, factor loadings for all manifest indicators of the latent constructs were significant in the expected direction ($p < .001$). Standardized loadings of the manifest indicators ranged from .43 to .95 ($M = .77$).

Because mediational tests often require that a significant association exists between the predictor and the outcome (Baron & Kenny, 1986; MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002), we first conducted an SEM, which constrained paths between attention difficulties (i.e., mediator) and school problems (i.e., outcome) at Waves 2 and 3 to 0. In this same model, we estimated the direct, predictive paths between latent constructs of Wave 1 insecure representations and (a) Waves 1 and 2 attention difficulties and (b) Waves 2 and 3 school adjustment problems. To provide conservative tests of prospective associations between interparental insecurity and child functioning, structural paths were also specified between interparental insecurity and school adjustment at Wave 3 after controlling for Wave 2 school adjustment through the specification of autoregressive paths. In support of the precondition for testing mediation, insecure representations of interparental relations predicted lower school functioning at Wave 2, $\beta = -.21$, $p < .01$, and Wave 3, $\beta = -.11$, $p < .05$.

Table 1
Means, Standard Deviations, and Intercorrelations of the Main Variables in the Primary Analyses

Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Insecure internal representations of interparental relations																
1. Emotional insecurity	3.06	0.76	—													
2. Relationship quality	3.08	0.79	**.83***	—												
3. Maternal competence	3.06	0.78	**.86***	**.81***	—											
4. Paternal competence	3.02	0.77	**.87***	**.79***	**.84***	—										
Child attention difficulties (Wave 1)																
5. Parent report	2.60	1.04	.24*	.20*	.22*	.18*	—									
Child attention difficulties (Wave 2)																
6. CPT omission error	54.06	12.85	.22*	.15*	.17*	.20*	.30*	—								
7. CPT standard error variability	57.43	8.95	.25*	.18*	.23*	.24*	.27*	**.60***	—							
8. Mother report	16.71	6.67	.21*	.11	.22*	.17*	.44*	**.23***	**.30***	—						
9. Father report	17.32	7.56	.17*	.11	.16*	.15*	.37*	**.18***	**.22***	.57*	—					
Child school adjustment (Wave 2)																
10. Prosocial peer behavior	17.18	3.41	−.22*	−.24*	−.23*	−.17*	−.14*	−.12	−.14	−.12	−.24*	—				
11. Behavioral competence	10.51	2.23	−.15*	−.16*	−.16*	−.12	−.26*	−.21*	−.20*	−.24*	−.31*	**.50***	—			
12. Cooperative participation	12.10	2.59	−.10	−.10	−.12	−.09	−.27*	−.27*	−.22*	−.21*	−.37*	**.54***	**.82***	—		
Child school adjustment (Wave 3)																
13. Prosocial peer behavior	17.01	3.16	−.21*	−.22*	−.25*	−.18*	−.31*	−.25*	−.21*	−.31*	−.24*	.39*	.43*	.46*	—	
14. Behavioral competence	10.54	2.23	−.22*	−.21*	−.24*	−.19*	−.37*	−.35*	−.25*	−.33*	−.28*	.40*	.75*	.68*	**.60***	—
15. Cooperative participation	11.91	2.86	−.16*	−.15*	−.19*	−.16*	−.39*	−.24*	−.21*	−.44*	−.32*	.26*	.59*	.52*	**.59***	**.75***

Note. Boldface denotes intercorrelations among manifest indicators of latent constructs. CPT = Continuous Performance Test.
**p* < .05.

Therefore, we subsequently tested a mediational model in which paths between attention difficulties and school adjustment were freely estimated. The resulting model, which is depicted in Figure 1, provided a good representation of the data, $\chi^2(80, N = 216) = 140.28$, $p < .001$, root mean square error of approximation = 0.06, and χ^2/df ratio = 1.75, comparative fit index = 0.97, and Tucker–Lewis index = 0.95. Inclusion of the structural paths between attention difficulties and school adjustment resulted in notable reductions in the magnitude of the path coefficients between insecure internal representations and school adjustment at Wave 2 (β reduced by 81% from −.21, $p < .01$, to −.04, *ns*) and Wave 3 (β reduced by 64% from −.11, $p < .05$, to −.04, *ns*). Consistent with hypotheses, analysis of the structural paths depicted in Figure 1 indicated that insecure representations at Wave 1 predicted attention difficulties at Wave 2, β = .24, $p < .01$, even after specifying the path between the Wave 1 proxy measure of attention difficulties and Wave 2 attention difficulties, β = .52, $p < .001$. Children's attention difficulties, in turn, were associated with poorer school adjustment at Wave 2, β = −.45, $p < .001$. In addition, even after estimating the autoregressive path between Waves 2 and 3 school adjustment (β = .70, $p < .001$), attention difficulty was a significant predictor of decreases in school adjustment at Wave 3, β = −.25, $p < .01$. In further support of the mediational role of attention difficulties, follow-up tests using MacKinnon et al.'s (2002) procedures for indirect effects indicated that the indirect pathways involving interparental insecurity and attention difficulties were significant in the prediction of child school adjustment at Wave 2, $z' = 2.43$, $p < .05$, and Wave 3, $z' = 2.15$, $p < .05$.

The role of child gender and research site in the mediational pathways. To further explore the generalizability of the mediational pathways, we examined

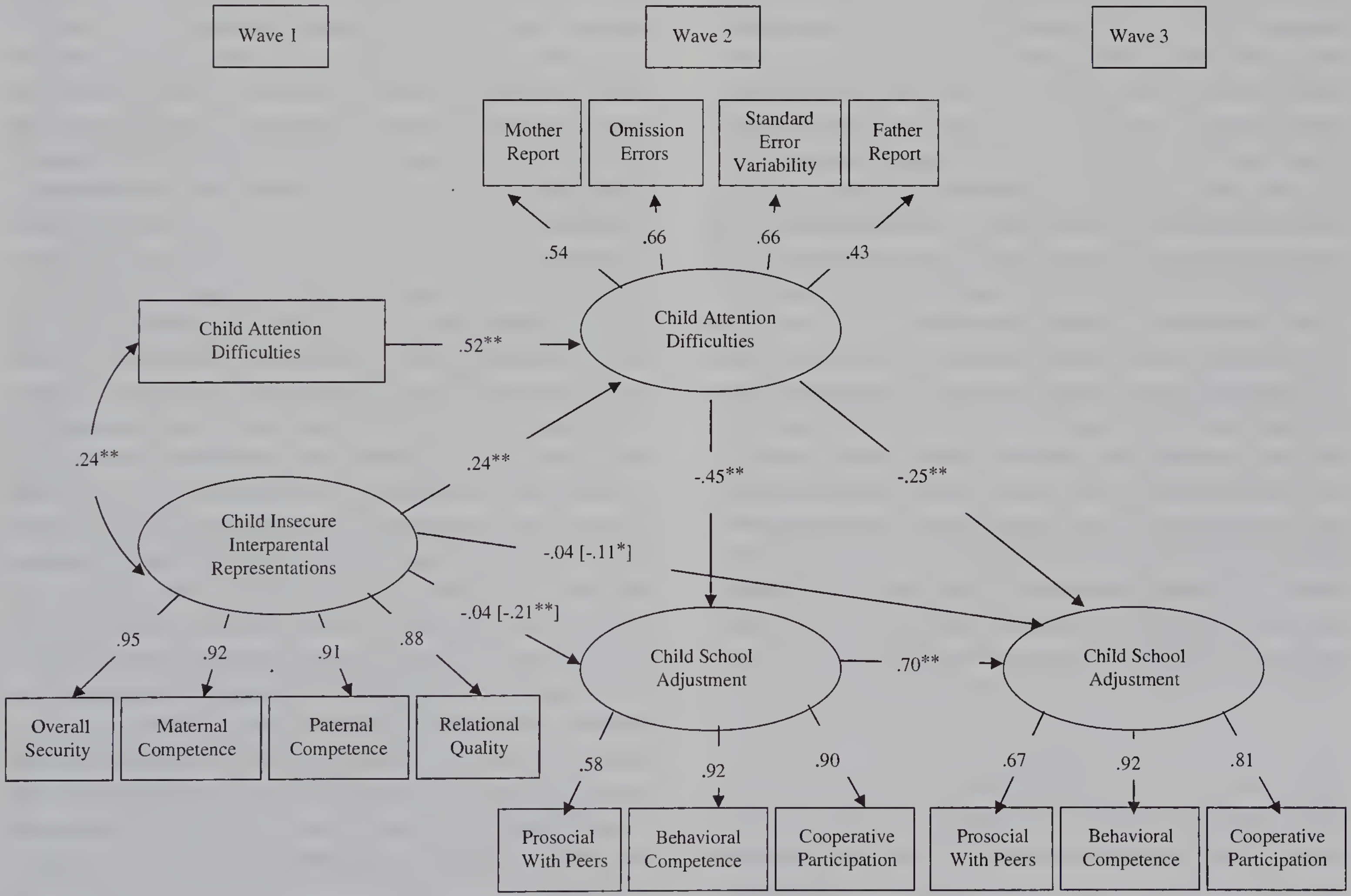

Wave 1
Wave 2
Wave 3
Mother Report
Omission Errors
Standard Error Variability
Father Report
.54
.66
.66
.43
Child Attention Difficulties
.52**
Child Attention Difficulties
.24**
.24**
-.45**
-.25**
Child Insecure Interparental Representations
-.04 [-.11*]
-.04 [-.21**]
.95
.92
.91
.88
Overall Security
Maternal Competence
Paternal Competence
Relational Quality
Child School Adjustment
.70**
Child School Adjustment
.58
.92
.90
.67
.92
.81
Prosocial With Peers
Behavioral Competence
Cooperative Participation
Prosocial With Peers
Behavioral Competence
Cooperative Participation

Figure 1. A structural equation model testing children's attention problems as a mediator of associations between their insecure internal representations of interparental relations and school adjustment. Parameter estimates for the structural paths are standardized path coefficients.

Note. Bracketed path coefficients reflect parameter estimates when constraining paths between the attention difficulties and the child school adjustment at Waves 2 and 3. Path coefficients without brackets denote parameter estimates for a model in which all pathways were freely estimated.

$^*p < .05$. $^{**}p < .01$.

whether inclusion of child gender or research site altered the associations among insecure internal representations, attention difficulties, and school adjustment as covariates or moderators. First, to examine the comparability of the findings with child gender and research site as covariates in the model, we included paths among child gender and research site and each of the latent constructs depicted in the model in Figure 1. Supporting the stability of the mediational findings, the results did not change with the addition of child gender and research site as predictors in the model.

Next, to test the moderating role of child gender and research site, we conducted successive multigroup models by splitting the data by child gender (i.e., boys vs. girls) and research site (i.e., NY vs. IN) for each mediational model. Analyses of whether different processes were operating in models across child gender involved comparing a model in which structural coefficients for the autoregressive and indirect paths were constrained to equivalence across boys and girls with a model in which all parameters were allowed to vary freely across gender. Likewise, corresponding comparisons of constrained and unconstrained models across research site were also conducted. To accommodate the smaller sample sizes resulting from the splitting of the data, it was necessary to conduct the multigroup path analyses using manifest variables for the primary constructs in the model (i.e., internal representations, attention difficulties, school adjustment at each time point). Manifest variables of each primary construct were calculated by standardizing their respective indicators and summing them to form a composite. Comparisons of the constrained and free-to-vary models revealed no difference in fit for either the child gender or the research site analyses, thereby indicating that child gender and research site did not moderate any of the proposed pathways. Thus, the mediational pathways in Figure 1 were statistically indistinguishable across child gender and research site.

Discussion

Although insecure representations of the interparental relationship have been shown to pose a risk for children's psychological adjustment (Cummings, Schermerhorn, Davies, Goeke-Morey, & Cummings, 2006), the processes accounting for the vulnerability experienced by children are little explored. Therefore, as a first step in testing a process-oriented model of the implications of emotional security, the objective was to examine children's attention difficulties as potential mediating mechanisms in the link between their insecure representations of interparental relations and psychological adjustment. According to EST, attention problems may develop from elevated concerns about security in the interparental relationship and, in turn, undermine children's functioning in the face of challenges arising in other settings (Davies, Winter, et al., 2006). Supporting this hypothesis, children's insecure representations of interparental relationships at Wave 1 predicted a multimethod, multiinformant assessment of children's attention difficulties 1 year later at Wave 2, even after controlling for Wave 1 assessment of attention difficulties. Attention difficulties, in turn, predicted concurrent school problems and increases in school problems over a 1-year period.

Interpreted within EST, multiple interlocking processes may account for why attention is a mediator of insecure internal representations. If internal representations serve as lenses for simplifying, interpreting, and responding to subsequent challenging contexts, then children who have negative interpretations of interparental relations may be primed to scan and interpret subsequent challenging settings for similar threats (Davies, Winter, et al., 2006; Johnston & Roseby, 1997). In accord with this premise, developmental neuroscience models stress that children who repeatedly experience family discord are disproportionately proficient at processing threatening interpersonal stimuli (e.g., Pollack & Tolley-Schell, 2004). However, this proficiency may occur at the expense of adeptly focusing and sustaining attention in nonthreatening settings (Derryberry & Tucker, 2006). Attention models further suggest that neurological restrictions limit capacities for successfully engaging in multiple, concurrent self-regulatory acts. Thus, representational tendencies to process negative stimuli may conflict with other demands on the attention network (Botvinick, Braver, Barch, Carter, & Cohen, 2001; Rothbart & Posner, 2006; Schimmack, 2005). The resulting cognitive interference may be manifested in disruptions in focusing and sustaining attention on relevant tasks. These difficulties, in turn, may be further expressed through problems in effectively encoding and processing interpersonal signals and communications essential for successfully adapting to social and academic challenges in school (Barkley et al., 2002).

The regulatory depletion model provides a complementary account of how children's insecure representations of interparental relations may give rise to attention difficulties. A primary assumption is that the regulation of responding in one domain of functioning (e.g., social cognitive) causes impairments in other response domains (e.g., cognitive problem-solving abilities) by depleting a shared reservoir of resources (Glass, Singer, & Friedman, 1969; Muraven, Tice, & Baumeister, 1998). By extension, difficulties preserving security may deplete a common pool of self-regulatory resources, thereby creating deficits in other domains of functioning. For example, the tendency for children with insecure representations to experience affective and behavioral dysregulation may create difficulties in persevering on challenging attention tasks. The depletion of resources accompanying insecurity may also engender impairments in inhibitory control, characterized by the tendency to forestall a dominant reaction in favor of a subdominant response. Within the context of attention processing, this may result in difficulties remaining engaged in academic tasks and complying with rules in the face of distractions and growing boredom (Barkley, 1997). Other pathways of resource depletion may also be evident. For example, research has shown that sleep disturbance is an intermediary mechanism linking insecurity in the interparental relationship and children's academic, cognitive, and emotional problems (El-Sheikh, Buckhalt, Cummings, & Keller, 2007; El-Sheikh, Buckhalt, Keller, Cummings, & Acebo, 2007). Thus, one possibility is that insecurity in the interparental relationship may engender attention difficulties in children by disrupting their sleep patterns.

In placing the results in broader perspective, as expected, attention did not fully mediate the link between insecure internal representations of interparental relations and school problems. Thus, consistent with recent formulations of EST (Davies, Winter, et al., 2006), other neurobiological, emotional, and cognitive processes may also act as operative mechanisms in producing the developmental consequences of emotional insecurity. For example, emotion dysregulation, poor inhibitory control, working memory impairments, and perceptual sensitivity to negative stimuli are not only associated with attention difficulties (e.g., Barkley, 1997). These same processes are also conceptualized in EST as mechanisms accounting for why children's concerns about security increase their risk for psychological problems (Davies & Sturge-Apple, 2007). Therefore, emotion dysregulation and executive functioning variables may also help to account for the link between insecure representations and children's adjustment difficulties. In addition, biopsychological models will likely serve as valuable sources of hypotheses about the neurobiological mechanisms (e.g., hypothalamic–pituitary–adrenal axis reactivity to threat, serotonergic system dysregulation) that may underlie the mediational role of attention difficulties (e.g., Van Goozen, Fairchild, Snoek, & Harold, 2007). Therefore, examining the role of attention impairments within a broader model of neuropsychological, biological, and self-regulatory problems is an important research direction.

Several limitations of our study are relevant in interpreting the results. First, although our longitudinal analysis of change is an advance over the predominant use of cross-sectional research on internal representations (for some exceptions, see Grych et al., 2003; Harold et al., 2004), the present results do not fully delineate directionality in relationships between children's insecure representations, attention difficulties, and their school adjustment. For example, even though the moderate to strong magnitude of the path between Waves 1 and 2 attention difficulties lends support for the validity of the Wave 1 inattention variable, the psychometric properties of the abbreviated Wave 1 measure may be weaker than the comparable Wave 2 assessment. As a result, the path between Wave 1 internal representations and Wave 2 attention difficulties may be inflated by our use of a proxy measure of inattention in the autoregressive path. However, with regard to specifying directionality, our primary aim was testing whether Wave 2 attention difficulties predicted subsequent change in school problems. In building our analytic model, we followed Cole and Maxwell's (2003) guidelines for authoritatively testing prospective mediational pathways. Therefore, for reasons of parsimony in already complex model tests and consistent with these guidelines for testing mediational pathways, some pathways (i.e., school outcomes at Wave 1) were not regarded as central components for the analyses.

Second, the autoregressive components of the analyses do not rule out the possible operation of other processes or extraneous variables in accounting for the mediational pathways. For example, it may be possible that the mediational effects of attention difficulties are simply rough markers for the more proximal, mediational role of coexisting neuropsychological and self-regulatory impairments (Barkley et al., 2002). Given the empirical documentation of bidirectional relations between children's adjustment problems and parenting difficulties (e.g., Lytton, 1990; Snyder, Cramer, Afrank, & Patterson, 2005), more complex reciprocal chains of processes may also be operating. For example, it is possible that school and attention difficulties may increase children's insecure representations by directly straining interparental relationships. Including broader assessments of children's responses to interparental conflict and their psychological adjustment in future research would also permit more comprehensive tests of the generalizability of the mediational pathways identified in this study. Thus, as with any study, the present findings should be considered within the larger context of developmental processes.

Third, although the sociodemographic characteristics of our sample were representative of the larger population of families within the Northeastern and Midwestern regions of the United States, our community sample consisted of predominantly White families. Therefore, the relationships documented in this study may not necessarily generalize to families with other racial or ethnic backgrounds or families who are facing considerable adversity. Suggestions regarding how our findings could differ across social and demographic conditions are highly speculative at this early stage of research. However, consideration of these factors raises interesting questions. For example, it is possible that African American and Hispanic children may be less susceptible to the deleterious effects of insecurity because extended family networks increase children's access to sources of support (McLoyd, Harper, & Cooper, 2001). In addition, exposure to multiple forms of adversity (e.g., poverty, violence, neuropsychological impairments) that are commonly found in higher risk samples may also be such powerful predictors of children's adjustment problems that they override the effects of insecurity (Davies, Winter, et al., 2006). Therefore, understanding how various contextual and demographic variables may alter the pathways of insecurity is an important research direction.

Despite these limitations, the longitudinal findings of this multimethod study support the hypothesis that difficulty in focusing and sustaining attention is a key mechanism in explaining why children's insecure representations of interparental relations increase their risk for experiencing problems in school. From a scientific perspective, this study represents the first attempt to identify possible mechanisms mediating children's emotional insecurity. As such, it may provide a blueprint for formulating and testing other possible mechanisms that mediate insecurity. In addition, although care should be exercised in applying our results to clinical practice, the findings may serve as guideposts in the development of clinical goals and tools for prevention and intervention programs targeting children from discordant homes. EST specifically proposes that insecure representations of family relationships reflect dispositions to identify and process threatening events in the family. In high-conflict homes, this insecure prototype of scanning the family context for danger has high adaptive value for children because it permits them quickly identify and cope with conflict before it becomes a bigger threat. Thus, implementing prevention and intervention programs designed to alter their insecure representations may be ineffective or even harmful for many children. Therefore, focusing on interrupting the pathogenic processes set in motion by children's concerns about security may be a feasible alternative to treatment programs designed to alter children's patterns of processing or responding to interparental conflict (Davies et al., 2006). The results in the present study specifically support the notion that children's attention difficulties is a key pathogenic process that accounts for the link between their insecure representations of the interparental relationship and their scholastic adjustment problems. Interpreted in this context, our mediational findings suggest that programs designed to optimize attention processes in social and academic settings may provide one method for protecting children from the negative sequelae associated with insecure representations.

References

Arbuckle, J. L. (2006). *Amos 7.0 user's guide.* Chicago: SPSS.

Barkley, R. A. (1997). Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of ADHD. *Psychological Bulletin, 121,* 65–94.

Barkley, R. A., Fischer, M., Smallish, L., & Fletcher, K. (2002). The persistence of attention-deficit/hyperactivity disorder into young adulthood as a function of reporting source and definition of disorder. *Journal of Abnormal Psychology, 111,* 279–289.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173–1182.

Birch, S. H., & Ladd, G. W. (1997). The teacher-child relationship and children's early school adjustment. *Journal of School Psychology, 35,* 61–79.

Boivin, M., & Begin, G. (1989). Peer status and self-perception among early elementary school children: The case of the rejected children. *Child Development, 60,* 591–596.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108,* 624–652.

Bretherton, I., Oppenheim, D., Buchsbaum, H., Emde, R. N., & the MacArthur Narrative Group (1990). *MacArthur Story Stem Battery.* Unpublished manuscript.

Campbell, S. B., Pierce, E. W., March, C. L., Ewing, L. J., & Szumowski, E. K. (1994). Hard-to-manage preschool boys: Symptomatic behavior across contexts and time. *Child Development, 65,* 836–851.

Carlson, E. A., Jacobvitz, D., & Sroufe, L. A. (1995). A developmental investigation of inattentiveness and hyperactivity. *Child Development, 66,* 37–54.

Cicchetti, D., Cummings, E. M., Greenberg, M. T., & Marvin, R. S. (1990). An organizational perspective on attachment beyond infancy: Implications for theory, measurement, and research. In M. T. Greenberg, D. Cicchetti, & E. M. Cummings (Eds.), *Attachment in the preschool years: Theory, research, and intervention* (pp. 3–49). Chicago: University of Chicago Press.

Clark, K. E., & Ladd, G. W. (2000). Connectedness and autonomy support in parent-child relationships: Links to children's socioemotional orientation and peer relationships. *Developmental Psychology, 36*, 485–498.

Cole, D. A., & Maxwell, S. E. (2003). Testing mediational models with longitudinal data: Questions and tips in the use of structural equation modeling. *Journal of Abnormal Psychology, 112*, 558–577.

Conners, C. K. (2000). *Conners' Continuous Perfomance Test II*. North Tonawanda, NY: Multi-Health Systems.

Cummings, E. M., & Davies, P. T. (1996). Emotional security as a regulatory process in normal development and the development of psychopathology. *Development and Psychopathology, 8*, 123–139.

Cummings, E. M., Schermerhorn, A. C., Davies, P. T., Goeke-Morey, M. C., & Cummings, J. S. (2006). Interparental discord and child adjustment: Prospective investigations of emotional security as an explanatory mechanism. *Child Development, 77*, 132–152.

Dadds, M. R., Atkinson, E., Turner, C., Blums, G. J., & Lendich, B. (1999). Family conflict and child adjustment: Evidence for a cognitive-contextual model of intergenerational transmission. *Journal of Family Psychology, 13*, 194–208.

Davies, P. T., & Cummings, E. M. (1994). Marital conflict and child adjustment: An emotional security hypothesis. *Psychological Bulletin, 116*, 387–411.

Davies, P. T., & Cummings, E. M. (1998). Exploring children's emotional security as a mediator of the link between marital relations and child adjustment. *Child Development, 69*, 124–139.

Davies, P. T., & Sturge-Apple, M. L. (2007). Advances in the formulation of emotional security theory: An ethologically-based perspective. *Advances in Child Behavior and Development, 35*, 87–137.

Davies, P. T., Sturge-Apple, M. L., Winter, M. A., Cummings, E. M., & Farrell, D. (2006). Child adaptational development in contexts of interparental conflict over time. *Child Development, 77*, 218–233.

Davies, P. T., Winter, M. A., & Cicchetti, D. (2006). The implications of emotional security theory for understanding and treating childhood psychopathology. *Development and Psychopathology, 18*, 707–735.

Derryberry, D., & Tucker, D. M. (2006). Motivation, self-regulation, and self-organization. In D. Cicchetti & D. J. Cohen (Eds.), *Developmental psychopathology: Vol. 2. Developmental neuroscience* (2nd ed., pp. 502–532). New York: Wiley.

Dunn, J., & Davies, L. (2001). Sibling relationships and interparental conflict. In J. H. Grych & F. D. Fincham (Eds.), *Interparental conflict and child development: Theory, research, and application* (pp. 273–290). Cambridge, UK: Cambridge University Press.

Eisenberg, N., Fabes, R. A., Bernzweig, J., Karbon, M., Poulin, R., & Hanish, L. (1993). The relations of emotionality and regulation to preschoolers' social skills and sociometric status. *Child Development, 64*, 1418–1438.

Eisenberg, N., Fabes, R. A., Guthrie, I. K., & Reiser, M. (2000). Dispositional emotionality and regulation: Their role in predicting quality of social functioning. *Journal of Personality and Social Psychology, 78*, 136–157.

Eisenberg, N., Guthrie, I. K., Fabes, R. A., Shepard, S., Losoya, S., Murphy, B. C., et al. (2000). Prediction of elementary school children's externalizing problem behaviors from attentional and behavioral regulation and negative emotionality. *Child Development, 71*, 1367–1382.

El-Sheikh, M., Buckhalt, J. A., Cummings, E. M., & Keller, P. (2007). Sleep disruptions and emotional insecurity are pathways of risk for children. *Journal of Child Psychology and Psychiatry, 48*, 88–96.

El-Sheikh, M., Buckhalt, J. A., Keller, P. S., Cummings, E. M., & Acebo, C. (2007). Child emotional insecurity and academic achievement: The role of sleep disruptions. *Journal of Family Psychology, 21*, 29–38.

Entwisle, D. R. (1995). The role of schools in sustaining early childhood program benefits. *Future of Children, 5*, 133–144.

Erel, O., & Burman, B. (1995). Interrelatedness of marital relations and parent-child relations: A meta-analytic review. *Psychological Bulletin, 118*, 108–132.

Fearon, R. M. P., & Belsky, J. (2004). Attachment and attention: Protection in relation to gender and cumulative social-contextual adversity. *Child Development, 75*, 1677–1693.

Glass, D. C., Singer, J. E., & Friedman, L. N. (1969). Psychic cost of adaptation to an environmental stressor. *Journal of Personality and Social Psychology, 12*, 200–210.

Gollob, H. F., & Reichardt, C. S. (1987). Taking account of time lags in causal models. *Child Development, 58*, 80–92.

Groot, A. S., de Sonneville, L. M. J., Stins, J. F., & Boomsa, D. I. (2004). Familial influences on sustained attention in preschoolers. *Journal of Child Psychology and Psychiatry, 45*, 306–314.

Grych, J. H., & Fincham, F. D. (1990). Marital conflict and children's adjustment: A cognitive-contextual framework. *Psychological Bulletin, 108*, 267–290.

Grych, J. H., & Fincham, F. D. (2001). Interparental conflict and child adjustment: An overview. In J. H. Grych & F. D. Fincham (Eds.), *Interparental conflict and child development: Theory, research, and application* (pp. 1–8). Cambridge, UK: Cambridge University Press.

Grych, J. H., Fincham, F. D., Jouriles, E. N., & McDonald, R. (2000). Interparental conflict and child adjustment: Testing the mediational role of appraisals in the cognitive-contextual framework. *Child Development, 71*, 1648–1661.

Grych, J. H., Harold, G. T., & Miles, C. J. (2003). A prospective investigation of appraisals as mediators of the link between interparental conflict and child adjustment. *Child Development*, *74*, 1176–1193.

Harold, G. T., Shelton, K. H., Goeke-Morey, M. C., & Cummings, E. M. (2004). Marital conflict, child emotional security about family relationships and child adjustment. *Social Development*, *13*, 350–376.

Harter, S., & Pike, R. (1984). The pictorial scale of perceived competence and social acceptance for young children. *Child Development*, *55*, 1969–1982.

Johnston, J. R., & Roseby, V. (1997). *In the name of the child: A developmental approach to understanding and helping children of conflicted and violent divorce*. New York: Free Press.

Kochenderfer, B. J., & Ladd, G. W. (1996). Peer victimization: Manifestations and relations to school adjustment in kindergarten. *Journal of School Psychology*, *34*, 267–283.

Ladd, G. W., & Burgess, K. B. (2001). Do relational risks and protective factors moderate the linkages between childhood aggression and early psychological and school adjustment? *Child Development*, *72*, 1579–1601.

Ladd, G. W., Herald, S. L., & Andrews, R. K. (2006). Young children's peer relations and social competence. In B. Spodek & O. N. Saracho (Eds.), *Handbook of research on the education of young children* (2nd ed., pp. 23–54). Mahwah, NJ: Erlbaum.

Ladd, G. W., & Profilet, S. M. (1996). The child behavior scale: A teacher-report measure of young children's aggressive, withdrawn, and prosocial behaviors. *Developmental Psychology*, *32*, 1008–1024.

Lemery, K. S., Essex, M. J., & Smider, N. A. (2002). Revealing the relation between temperament and behavior problem symptoms by eliminating measurement confounding: Expert ratings and factor analyses. *Child Development*, *73*, 867–882.

Loeber, R., Green, S. M., Lahey, B. B., Christ, M. A., & Frick, P. J. (1992). Developmental sequences in the age of onset of disruptive child behaviors. *Journal of Child and Family Studies*, *1*, 21–41.

Lytton, H. (1990). Child and parent effects in boys' conduct disorder: A reinterpretation. *Developmental Psychology*, *26*, 683–697.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods*, *7*, 83–104.

McLoyd, V. C., Harper, C. I., & Cooper, N. L. (2001). Ethnic minority status, interparental conflict, and child adjustment. In J. H. Grych & F. D. Fincham (Eds.), *Interparental conflict and child development: Theory, research, and application* (pp. 98–125). Cambridge, UK: Cambridge University Press.

Muraven, M., Tice, D. M., & Baumeister, R. F. (1998). Self-control as a limited resource: Regulatory depletion patterns. *Journal of Personality and Social Psychology*, *74*, 774–789.

National Institute of Child Health and Human Development Early Child Care Network. (2003). Do children's attention processes mediate the link between family predictors and school readiness? *Developmental Psychology*, *39*, 581–593.

National Institute of Child Health and Human Development Early Child Care Network. (2005). Predicting individual differences in attention, memory, and planning in first graders from experiences at home, child care, and school. *Developmental Psychology*, *41*, 99–114.

Pollack, S. D., & Tolley-Schell, S. (2004). Attention, emotion, and the development of psychopathology. In M. I. Posner (Ed.), *Cognitive neuroscience of attention* (pp. 357–368). New York: Gilford.

Putnam, S. P., & Rothbart, M. K. (2006). Development of short and very short forms of the children's behavior questionnaire. *Journal of Personality Assessment*, *87*, 102–112.

Reeves, C. B., Palmer, S. L., Reddick, W. E., Merchant, T. E., Buchanan, G. M., Gajjar, A., et al. (2006). Attention and memory functioning among pediatric patients with medulloblastoma. *Journal of Pediatric Psychology*, *31*, 272–280.

Rothbart, M. K., Ahadi, S. A., Hersey, K. L., & Fisher, P. (2001). Investigations of temperament at three to seven years: The children's behavior questionnaire. *Child Development*, *72*, 1394–1408.

Rothbart, M. K., & Posner, M. I. (2006). Temperament, attention, and developmental psychopathology. Motivation, self-regulation, and self-organization. In D. Cicchetti & D. J. Cohen (Eds.), *Developmental psychopathology: Vol. 2. Developmental neuroscience* (2nd ed., pp. 465–501). New York: Wiley.

Schimmack, U. (2005). Attentional interference effects of emotional pictures: Threat, negativity, or arousal? *Emotion*, *5*, 55–66.

Seidel, W. T., & Joschko, M. (1990). Evidence of difficulties in sustained attention in children with ADHD. *Journal of Abnormal Child Psychology*, *18*, 217–229.

Snyder, J., Cramer, A., Afrank, J., & Patterson, G. R. (2005). The contributions of ineffective discipline and parental hostile attributions of child misbehavior to the development of conduct problems at home and school. *Developmental Psychology*, *41*, 30–41.

Van Goozen, S. H. M., Fairchild, G., Snoek, H., & Harold, G. T. (2007). The evidence for a neurobiological model of childhood antisocial behavior. *Psychological Bulletin*, *133*, 149–182.

Webster-Stratton, C., & Lindsay, D. W. (1999). Social competence and conduct problems in young children: Issues in assessment. *Journal of Clinical Child Psychology*, *28*, 25–43.

Zelkowitz, P., Paris, J., Guzder, J., & Feldman, R. (2001). Diatheses and stressors in borderline pathology of childhood: The role of neuropsychological risk and trauma. *Journal of the American Academy of Child and Adolescent Psychiatry*, *40*, 100–105.

Child Development, September/October 2008, Volume 79, Number 5, Pages 1583–1593

Developmental Trajectories of Sex-Typed Behavior in Boys and Girls: A Longitudinal General Population Study of Children Aged 2.5–8 Years

Susan Golombok, John Rust,
Karyofyllis Zervoulis, and Tim Croudace
University of Cambridge

Jean Golding
University of Bristol

Melissa Hines
University of Cambridge

The stability of sex-typed behavior from the preschool to the middle school years was examined. The Preschool Activities Inventory, a measure of within-sex variation in sex-typed behavior, was completed by the primary caregiver when the child was 2.5, 3.5, and 5 years, and a modified version, the Child Activities Inventory, was completed by the child at age 8. The investigation involved a general population sample of 2,726 boys and 2,775 girls. Sex-typed behavior increased through the preschool years, and those children who were the most sex typed at age 2.5 were still the most sex typed at age 5, with those children who showed the highest levels of sex-typed behavior during the preschool years continuing to do so at age 8.

From an early age, boys and girls differ in their behavior (for a review, see Ruble, Martin, & Berenbaum, 2006). This does not mean that all boys engage in typically male behaviors, or that all girls engage in typically female behaviors, all the time. There is a great deal of overlap between the sexes with some girls being more "boyish" than the average boy and some boys behaving in ways that are more typical of a "girlish" girl. One of the most striking differences between boys and girls is their preference for different types of toys. This can be detected in 1-year-olds (Snow, Jacklin, & Maccoby, 1983), and by 3 years of age, girls are much more likely than boys to play with dolls, dolls' houses, tea sets, and other domestic toys, whereas boys are more often found with toy guns, swords, cars, trains, and trucks (Dunn & Hughes, 2001; O'Brien & Huston, 1985; Servin, Bohlin, & Berlin, 1999; Sutton-Smith & Rosenberg, 1971). This sex difference in toy preferences continues to characterize the preschool and early school years. For example, in a study of 3- to 5-year-olds, a greater preference was demonstrated among boys for a car, an airplane, and a tool set, and among girls for a tea set, a doll, and a hair care set (Martin, Wood, & Little, 1990). Children in this study were also more likely to report that they preferred unfamiliar neutral items such as a pizza cutter and a hole puncher if presented to them as something that children of their sex really like.

In addition, sex differences in playmate preference become apparent in the preschool years. From as early as 3 years of age, girls prefer other girls as playmates and boys prefer to play with boys (Maccoby & Jacklin, 1987). This phenomenon, known as gender segregation, is not specific to particular nationalities or cultures. Children's preference for same-sex playmates is a universal aspect of growing up (Whiting & Edwards, 1988). Gender segregation is most likely to occur when children are not with adults, especially when they are with others of a similar age. In a study of children in a day-care setting, the tendency for girls to play together was observed at age 2 (LaFreniere, Strayor, & Gauthier, 1984). For boys, the preference for male playmates began slightly later but was clearly established at 3 years of age. Howes (1988) reported

---

We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole Avon Longitudinal Study of Parents and Children (ALSPAC) team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, and nurses. The UK Medical Research Council, the Wellcome Trust, and the University of Bristol provide core support for ALSPAC. This research was specifically funded by the Wellcome Trust.

Correspondence concerning this article should be addressed to Susan Golombok, Centre for Family Research, Faculty of Social and Political Sciences, Free School Lane, University of Cambridge, Cambridge CB2 3RF, United Kingdom. Electronic mail may be sent to seg42@cam.ac.uk.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7905-0024

similar findings. Like LaFreniere et al., she observed children in day care and showed that 3-year-olds were more likely to form new friendships with children of the same sex. Gender segregation is also an important feature of the early school years. Maccoby and Jacklin (1987) examined the playmate preferences of 100 children at 4.5 years of age and again at 6.5 years. When first observed, the children were spending nearly 3 times as much time playing with same-sex peers than other-sex peers. By the second observation, only 2 years later, the amount of time spent playing with same-sex peers was more than 10 times greater than that spent with peers of the other sex. A similar increase in the preference for same-sex peers was demonstrated in a comparison between children in kindergarten and children in the early school years by Serbin, Powlishta, and Gulko (1993). The process of gender segregation involves not only the preference for same-sex playmates but also the avoidance of playmates of the other sex.

Differences in the play styles of boys and girls can also be seen from 3 years of age (Maccoby, 1998; Maccoby & Jacklin, 1987; Pitcher & Shultz, 1983). Boys tend to play in a more active, rough-and-tumble and sometimes physically aggressive, fashion than girls, who tend to talk more to each other and be more nurturant than boys. When girls are aggressive, this is more likely to take the form of behaviors intended to damage relationships such as exclusion from a circle of friends (Crick, Casas, & Mosher, 1997; McNeilly-Choque, Hart, Robinson, Nelson, & Olsen, 1996). Boys also like to play outdoors in large groups, whereas girls are more often to be found in twos and threes indoors. Differences in the way in which boys and girls resolve conflict are also apparent from age 3. Whereas girls incline toward reaching a compromise, it is more common for boys to use physical force (Sheldon, 1990). In addition, pretend play differentiates the sexes, with boys acting out heroic roles involving fighting and adventure and girls preferring to be family characters or dressing up in feminine clothes. As children move from preschool to the early school years, the play styles of boys and girls continue to diverge. Achieving dominance appears to be of particular importance to boys. In order to have status, boys need to be seen as "tough" (Maccoby, 1998). The way in which boys and girls communicate is also different. Girls talk to each other to form and strengthen relationships. Boys use language to give information, assert themselves, and command attention (Lever, 1976; Maccoby & Jacklin, 1987). Boys of this age like to play in large groups of other boys, whereas girls prefer the company of one or two female friends (Fabes, Martin, & Hanish, 2003; Maccoby, 2002). The nature of these relationships also differs between the sexes. Whereas girls' friendships are characterized by emotional and physical closeness, the friendships of boys are founded on shared activities and interests (Maccoby, 1998).

Although there is much evidence from comparisons between the sexes to show that group differences in behavior persist from early to middle childhood, little is known about the extent to which within-sex differences remain stable over time, that is, whether girls or boys who show very feminine, or very masculine, behavior at an early age continue to do so as they progress from the preschool to the middle school years (Ruble et al., 2006). One exception is Martin and Fabes's (2001) observational study of preschool children that demonstrated stability in sex-segregated play for boys and girls over a period of 6 months. A difficulty in addressing this question has been the lack of an assessment instrument designed to measure within-sex variation in gender role behavior that is suitable for use with large samples. A number of questionnaires have been developed to examine gender development in young children (see Beere, 1990, for a review). However, these have focused on between-sex, rather than within-sex, differences. An exception is the Preschool Activities Inventory (PSAI), a psychometrically constructed questionnaire specifically designed to differentiate "masculine" from "feminine" boys, and "masculine" from "feminine" girls, within a normal population sample, that is, to differentiate within, as well as between, the sexes. During the construction of the PSAI, items that, while discriminating between the sexes, failed to differentiate between masculine and feminine boys, or between masculine and feminine girls, were excluded. For more details of the construction of the PSAI, see Golombok and Rust (1993a, 1993b, 2007). Thus, the PSAI is ideally suited to the investigation of stability and change in the gender development of boys and girls. The present study became possible through collaboration with the Avon Longitudinal Study of Parents and Children, (ALSPAC), the first general population data set to include an assessment of childhood gender role behavior. The PSAI was administered when the children were aged 2.5, 3.5, and 5 years old, and an age-appropriate modified version, the Children's Activities Inventory (CAI), was administered at age 8. The aims of the present investigation were to examine (a) the stability of gender role behavior from the preschool to the middle school years, that is, to establish whether "masculine" boys and "feminine" girls, and "feminine" boys and "masculine" girls, continue to show sex-typed and cross-sex-typed behavior, respectively, as they grow up; (b) change in gender role behavior from the preschool to the middle school years, that is, to establish whether boys and/or girls become more or less sex typed as they grow up; and (c) whether "masculine" boys and "feminine" girls show more or less change in sex-typed behavior than "feminine" boys and "masculine" girls, that is, to establish whether sex-typed children show greater change in sex-typed behavior as they grow up than their cross-sex-typed counterparts.

A range of theories have been put forward to explain the processes involved in children's acquisition of sex-typed behavior, the most influential of which have arisen from biological, that is, early hormonal (Hines, 2004), social learning (Bandura, 1977; Mischel, 1966, 1970), and cognitive developmental (Kohlberg, 1966; Martin, 1993) perspectives. It is now generally agreed that sex differences in behavior result from an interplay among biological, psychological, and social mechanisms from early fetal development onward (Bussey & Bandura, 1999; Golombok & Fivush, 1994; Ruble et al., 2006). These perspectives would each predict stability in sex-typed behavior from ages 2.5 to 8 years old as biological influences would be expected to be maintained postnatally, the processes of differential reinforcement and modeling that characterize the classic social learning approach are apparent and are probably relatively stable throughout this age range, and the cognitive processes of gender concept formation and gender schematization occur during the preschool and middle school years. In addition, cognitive developmental explanations of gender development would predict an increase in sex-typed behavior as children grow older as it is argued that children must progress from the stage of basic gender identity through gender stability to reach the stage of gender constancy at around age 5 years and that sex-typed behavior becomes more pronounced as genetic schemas are elaborated. It is important to differentiate between sex-typed behavior and gender stereotyping, the latter of which becomes more flexible at ages 7–8 years (Ruble et al., 2006). To the extent that the processes described by the different theoretical perspectives work together in the development of sex-typed behavior, with inborn hormonal mechanisms forming the basis on which postnatal social and cognitive processes build, it may be expected that sex-typed behavior would not simply remain stable but would increase between the preschool and the middle school years. Social learning and cognitive perspectives might also predict that children who are highly sex typed as toddlers would become even more sex typed over time as these children may receive greater exposure to, and may be more likely to seek out, social and psychological influences that promote sex-typical development, such as interaction with same-sex peers, as they grow up.

Method

Participants

The sample was obtained from the ALSPAC, a geographic population study of almost 14,000 mothers and their children beginning in pregnancy (Golding, Pembrey, Jones, & the ALSPAC Study Team, 2001). The study enrolled all pregnant women who were resident within Avon, a clearly defined area of southwest England, who had an expected delivery date between April 1, 1991, and December 21, 1992. It was estimated that approximately 90% of the eligible population took part. The study area has a population of 1 million comprising the city of Bristol, with a population of 0.5 million, moderate-sized towns, and rural areas. The demographic characteristics of families in the study are closely comparable to those of families in the United Kingdom as a whole with respect to the type of area in which they live, the educational level of the parents, housing, and mobility (Baker, Morris, & Taylor, 1997). The children in the study are similar to children in the rest of the country with respect to the prevalence of preterm delivery, low birth weight, physical and mental disability, physical illness, and psychological disorder.

As part of the ALSPAC study, all mothers or primary caregivers were asked to complete the PSAI (Golombok & Rust, 1993a, 1993b, 2007) when the ALSPAC child was aged 2.5, 3.5, and 5 years old. The children themselves were administered the CAI (Golombok & Rust, 2007) at 8 years old. The present investigation involved 5,501 children (2,726 boys and 2,775 girls) for whom data were obtained at all four time points.

Measures

The PSAI is a questionnaire measure for which a parent or the child's primary caregiver indicates their child's involvement in a variety of sex-typed behaviors. The PSAI comprises 24 items, 12 masculine and 12 feminine, and is divided into three sections: *toys* (7 items), *activities* (11 items), and *characteristics* (6 items). Each section has its own stem. For toys, this is "Please answer the questions according to how often the child played with the following toys during the past month" and an example item is "Jewelry."

For activities, the stem is "Please answer the questions according to how often the child engaged in the following activities during the past month" and an example item is "Fighting." For characteristics, the stem "Please answer the questions according to how often the child showed the following characteristics" is used, and an example item is "Interest in snakes, spiders or insects." Each item is scored on a 5-point scale ranging from *never* through *hardly ever*, *sometimes*, and *often* to *very often*. The feminine items are reversed so that higher total scores represent more masculine-typical behavior and lower total scores represent more feminine-typical behavior. Thus, the higher the total score, the more masculine the behavior for both boys and girls. A benefit of reversing half of the scores in this way is that the effects of acquiescence, that is, the tendency of respondents to agree or disagree with an item regardless of its content, and of differences in level of overall activity, are balanced. The measure has been standardized on more than 2,000 children in the United Kingdom, the Netherlands, and the United States (Golombok & Rust, 1993a). Mean standardized scores on the inventory did not differ for children from these three countries and for the entire group were 61.66 (*SD* = 9.40) for boys and 38.72 (*SD* = 9.66) for girls. The standardization target was a mean of 40 (*SD* = 10) for girls and a mean of 60 (*SD* = 10) for boys. Split-half reliability was .66 for boys and .80 for girls, and test–retest reliability over a 1-year period was .62 for boys and .66 for girls. The inventory has been validated by comparing parental ratings to teacher ratings for children attending five different day-care centers. For boys, the correlation between parent and teacher ratings was .37, and for girls the correlation was .48, showing the inventory to be a valid measure of gender role.

The CAI is an adaptation of the PSAI for children aged 7–8 years old. Unlike the PSAI, the CAI is completed by the child rather than the parent or primary caregiver. Instead of giving the measure as a written questionnaire, the CAI is administered as a task using Harter's format as described below. The PSAI items were transformed to the Harter format in a pilot study, some additional age-appropriate items were included, and the instrument was shortened to 16 items on the basis of item analysis. The child is presented with two postboxes, one blue and one red, with two slots in each. The researcher then reads out a statement written on an envelope in blue ink (e.g., "Some children play with jewelry") while patting the blue postbox, followed by a statement written on a separate envelope in red ink (e.g., "But other children don't play with jewelry") while patting the red postbox. The child is then asked which group of children she/he feels most similar to. If she/he replies that she/he feels similar to the children who play with jewelry, she/he is asked to put the envelope with the blue writing in the blue postbox. If she/he replies that she/he feels similar to the children who do not play with jewelry, she/he is asked to put the envelope with the red writing in the red postbox. The child also has to choose which of the two slots to post the letter into by deciding if the statement is "really true" or "sort of true" of him or her. The researcher points to each slot in the postbox, labeled accordingly, while explaining this to the child. The child then posts the envelope into the slot that represents his or her response. Like the PSAI, the CAI was standardized to give a mean score of 40 (*SD* = 10) for girls and a mean score of 60 (*SD* = 10) for boys, thus allowing the means for the PSAI and the CAI to be compared. Split-half reliability for the CAI is .64 for both boys and girls.

Results

The means and standard deviations of the PSAI at ages 2.5, 3.5, and 5 years, and for the CAI at age 8 years, are presented in Table 1, showing that the instrument differentiates between boys and girls. As the distribution of both PSAI and CAI scores for the sexes combined is bimodal, the analyses were conducted for boys and girls separately unless stated otherwise. The split-half reliabilities also appear in Table 1.

Stability (test–retest reliability) was examined between all three PSAI measures (ages 2.5, 3.5, and 5 years). Stability coefficients for the PSAI between 2.5 years and 3.5 years, 2.5 years and 5 years, and 3.5 years and 5 years were .65, .52, and .68, respectively, for boys, and .63, .53, and .70, respectively, for girls, demonstrating high stability over time. All these correlations were significant at the 0.001 level.

As the measures were repeated over more than two time points, a multilevel modeling approach (MLM) was used to examine the nature of change over time. Multilevel models are models for data that are organized hierarchically and is a generic term for all models for hierarchical or nested data. In the present case, it refers to the nested structure of repeated measurements on individuals over time. Multilevel models yield the same results whether from a hierarchical linear or latent growth perspective. A series of MLM analyses were carried out using the SPSS mixed models program. MLM using this procedure enables the testing of two questions (a) On average, is there a change in sex-typed behavior over time? (b) Are there differences between children in change in sex-typed behavior over time (i.e., do some children

Table 1

Descriptive Statistics of Preschool Activities Inventory Scores and of Children's Activity Inventory Scores (Mean, SD, and Split-Half Reliability)

	Preschool Activities Inventory (30 months)		Preschool Activities Inventory (42 months)		Preschool Activities Inventory (57 months)		Children's Activities Inventory (8 years)	
	M (*SD*)	Split half	*M* (*SD*)	Split half	*M* (*SD*)	Split half	*M* (*SD*)	Split half
Boys (2,726)	60.23 (8.33)	.73	62.07 (8.78)	.77	64.18 (8.77)	.80	59.72 (11.48)	.64
Girls (2,775)	40.90 (8.31)	.73	37.04 (9.81)	.80	35.11 (9.41)	.84	39.95 (12.35)	.64

change more than other children)? A further advantage of this approach is that it generates maximum likelihood estimates of trajectories of sex-typed behavior over time for each child, and these can be correlated with other variables to examine the relationship between the trajectories and the extraneous variables.

Using the full data set of boys and girls combined, a trajectory analysis model was fitted to establish (a) the extent of the sex difference in PSAI scores and (b) the extent of the sex difference in the overall change in PSAI scores over time, that is, the difference between boys and girls in change over time. The model fitted was:

$$\mathrm{PSAI}_{ti} = \gamma_{00} + \gamma_{01}\mathrm{Sex}_i + \gamma_{10}\mathrm{Time}_{ti} + \mathrm{U}_{1i}\mathrm{Time}_{ti} \times \mathrm{Sex}_i + \mathrm{U}_{1i}\mathrm{Time}_{ti} + \mathrm{U}_{0i} + e_{ti}$$

Where PSAI_{ti} is the PSAI score of Child 1 at time t, γ_{00} = is the grand intercept, $\gamma_{01}\mathrm{Sex}_i$ represents between-sex differences, $\gamma_{10}\mathrm{Time}_{ti}$ is the grand slope, $\mathrm{U}_{1i}\mathrm{Time}_{ti} \times \mathrm{Sex}_i$ represents sex differences in trends, $\mathrm{U}_{1i}\mathrm{Time}_{ti}$ and U_{0i} represent repeated measures, and e_{ti} is the within-subject error term.

Results showed the expected finding that PSAI scores differed between the sexes, $t(5{,}441) = 89.25$, $p < .001$. The overall mean for boys was 62.15 ($N = 2{,}705$), and for girls, the overall mean was 37.68 ($N = 2{,}738$). In addition, there was a significant overall trend in PSAI scores over time for boys and girls combined, $t(5{,}442) = 38.16$, $p < .001$. The mean trend was −0.47 ($N = 5{,}443$) indicating a tendency for children overall to show less male sex-typed behavior (i.e., more female sex-typed behavior) as they grow older. However, this trend differed significantly between the sexes, $t(5{,}441) = 42.99$, $p < .001$, with a mean for boys of 1.98 ($N = 2{,}705$) and a mean for girls of −2.89 ($N = 2{,}738$), demonstrating that the trends for boys and girls were markedly different. Having established that boys and girls showed different means and trends in PSAI scores, MLM analyses were conducted separately for each sex.

The model fitted separately for boys and girls was:

$$\mathrm{PSAI}_{ti} = \gamma_{00} + \gamma_{10}\mathrm{Time}_{ti} + \mathrm{U}_{0i} + \mathrm{U}_{1i}\mathrm{Time}_{ti} + e_{ti}$$

Where PSAI_{ti} is the PSAI score of Child 1 at time t, γ_{00} = is the grand intercept, $\gamma_{10}\mathrm{Time}_{ti}$ is the grand slope, $\mathrm{U}_{1i}\mathrm{Time}_{ti}$ represents repeated measures, and e_{ti} is the within-subject error term.

For boys, the analysis gave a −2 restricted log likelihood of 54,838.9, with an average PSAI score of 60.18 at age 2.5, and a statistically significant average trend of 1.98, $t(2{,}704) = 24.57$, $p < .001$, indicating an increase in male sex-typed behavior over time. A log likelihood is the standard statistic for fitting maximum likelihood in multilevel models, although here we are not reporting the likelihood of alternative models of fit for these data. Estimation of the covariance parameters also showed a significant variation among boys in trends over time (Wald $Z = 11.55$, $p < .001$), showing that the sex-typed behavior of some boys changed more than others. For girls, the −2 restricted log likelihood was 56,179.9, with an average PSAI score of 40.58 at age 2.5, and a statistically significant average trend of −2.89, $t = (2{,}737) = -34.83$, $p < .001$, indicating an increase in female sex-typed behavior over time. Estimation of the covariance parameters showed a significant variation among girls in trends over time (Wald $Z = 10.75$, $p < .001$), showing that the sex-typed behavior of some girls changed more than others. To test the hypothesis that the degree of change in sex-typed behavior differed between the sexes, the trajectory scores derived for each child were reversed for girls (to compare the degree of sex typing rather than the degree of change) and a t test carried out, $t(5{,}442) = -21.53$, $p < .001$, indicating that girls showed a greater change in the direction of sex-typical behavior between the ages of 2.5 and 5 years than did boys.

In order to examine the nature of the relationship between PSAI scores and PSAI trajectories for boys and girls, PSAI scores at age 3.5 were binned (put into categories) at PSAI score intervals of 5 to generate nine categories of PSAI scores for boys and girls separately. PSAI at age 3.5, representing the middle value, that is, the pivotal point, was used rather the mean of the three PSAI scores, in order to retain the statistical parameters of the MLM model. One-way analyses of variance were then carried out with the PSAI trajectory as the dependent variable and the PSAI bin as the independent variable, $F(8, 2696) = 11.25$, $p < .001$, for boys, and $F(8, 2729) = 65.57$, $p < .001$, for girls. Polynomial contrasts were then carried out. For both boys and girls, statistically significant linear effects were found: boys, $F(1, 2696) = 61.04$, $p < .001$; girls, $F(1, 2729) = 407.92$, $p < .001$. Boys with high (i.e., more masculine) PSAI scores at age 3.5 showed significantly more change over time than boys with low (i.e., less masculine) PSAI scores (see Table 2). Similarly, girls with low (i.e., more feminine) PSAI scores at age 3.5 changed significantly more over time than girls with high (i.e., less feminine) PSAI scores (see Table 3). Thus, the children who were most sex typed to start with became even more so over time. Means and standard deviations of the trajectory scores at the various levels of PSAI scores at age 3.5 appear in Figures 1 and 2 for boys and girls, respectively. The relationship between the PSAI at age 3.5 and the CAI at age 8 was also examined for boys and girls separately. One-way analyses of variance were carried out with the CAI as the dependent variable and the PSAI bin as the independent variable. Significant effects were found for both boys, $F(8, 2696) = 7.08$, $p < .001$, and girls, $F(8, 2729) = 16.08$, $p < .001$. Polynomial contrasts were applied and statistically significant linear effects were found for boys, $F(1, 2696) = 31.52$, $p < .001$, and girls, $F(1, 2729) = 89.92$, $p < .001$. Boys with high (i.e., more masculine) PSAI scores at age 3.5 showed significantly higher CAI scores at age 8 than boys with low (i.e., less masculine) PSAI scores at age 3.5 (see Table 4 and Figure 3). Similarly, girls with low (i.e., more feminine) PSAI scores at age 3.5 showed lower CAI scores at age 8 than girls with high (i.e., less feminine) PSAI scores at age 3.5 (see Table 5 and Figure 4). Thus, the children who were most sex typed at age 3.5 continued to be so at age 8.

In order to examine the nature of the relationship between PSAI trajectory scores and CAI scores for boys and girls, that is, whether the rate of change in PSAI scores predicted CAI scores at age 8, PSAI trajectory scores were binned at intervals of 1.00 to generate eight categories of PSAI trajectory scores for boys and girls separately. One-way analyses of variance were then carried out with the CAI score as the dependent variable and the PSAI trajectory bin as the independent variable. Significant effects were found for both boys, $F(7, 2697) = 3.29$, $p < .002$, and girls, $F(7,$

Table 3

Means and Standard Deviations of Preschool Activities Inventory Trajectory Scores at 5-Point Intervals of Preschool Activities Inventory Score at Age 3.5 for Girls

Preschool Activities Inventory category	*n*	*M*	*SD*
≤ 19	86	−4.22	1.53
20–25	142	−3.84	1.33
25–30	394	−3.41	1.35
30–35	462	−3.38	1.43
35–40	608	−2.81	1.41
40–45	568	−2.60	1.42
45–50	264	−2.20	1.43
50–55	155	−1.87	1.75
> 55	59	−0.82	1.87
Total	2,738	−2.89	1.58

Table 2

Means and Standard Deviations of Preschool Activities Inventory Trajectory Scores at 5-Point Intervals of Preschool Activities Inventory Score at Age 3.5 for Boys

Preschool Activities Inventory Category	*n*	*M*	*SD*
≤ 44	69	1.38	1.89
45–50	152	1.66	1.31
50–55	394	1.64	1.56
55–60	486	1.78	1.47
60–65	628	2.08	1.47
65–70	443	2.18	1.59
70–75	357	2.10	1.66
75–80	115	2.55	1.59
> 80	61	2.82	1.61
Total	2,705	1.98	1.56

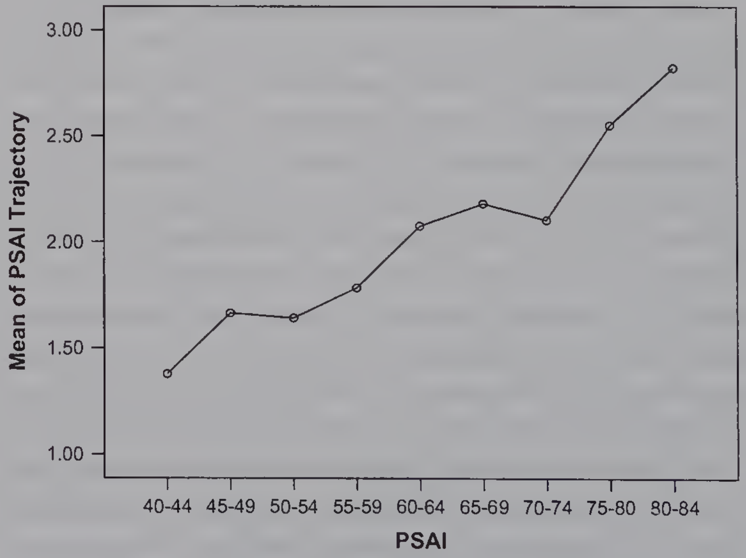


Figure 1. Mean of Preschool Activities Inventory (PSAI) trajectory scores plotted against PSAI scores at age 3.5 for boys.

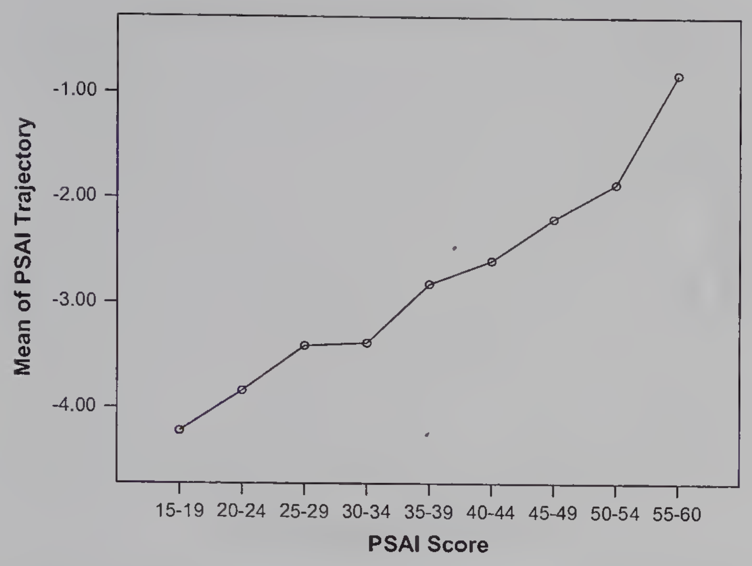


Figure 2. Mean of Preschool Activities Inventory (PSAI) trajectory scores plotted against PSAI scores at age 3.5 for girls.

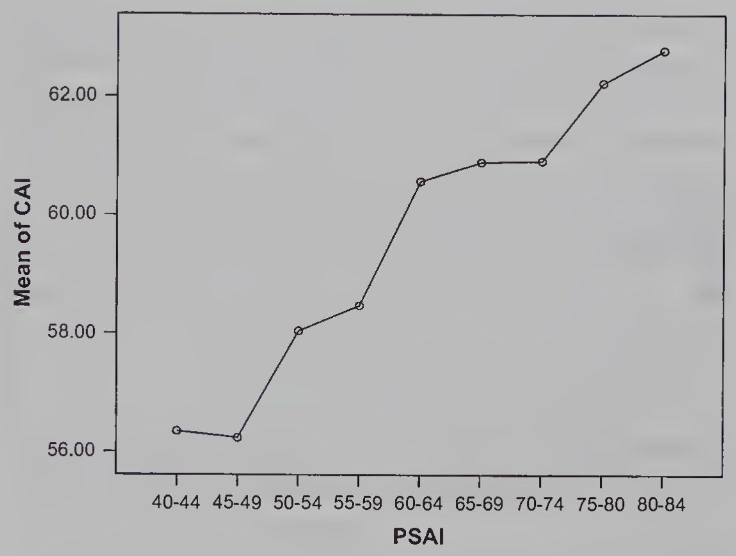


Figure 3. Mean of Children's Activities Inventory (CAI) scores plotted against Preschool Activities Inventory (PSAI) scores at age 3.5 for boys.

2730) = 17.63, $p < .001$. Polynomial contrasts were then carried out. Statistically significant linear effects were found for boys, $F(1, 2697) = 15.01, p < .001$, and girls, $F(1, 2730) = 77.30, p < .001$. Boys with higher PSAI trajectory scores (i.e., those who showed the greatest increase in masculinity) showed significantly higher levels of male sex-typed behavior at age 8 than boys who showed relatively less increase (see Table 6 and Figure 5). Similarly, girls with lower PSAI trajectory scores (i.e., those who showed the greatest increase in femininity) showed significantly higher levels of female sex-typed behavior at age 8 than girls who showed relatively less decrease (see Table 7 and Figure 6). For girls, but not for boys, a significant quadratic effect was also found, $F(1, 2730) = 13.10, p < .001$, indicating that girls who showed no increase in sex-typed behavior between the ages of 2.5 and 5 years showed the least feminine sex-typed behavior at age 8. These girls obtained CAI scores at age 8 that fell halfway between the average score for girls and the average score for boys. Thus, the children who showed the greatest increase in sex typing from 2.5 to 5 years old were the most sex typed at age 8, and the girls who showed no change in sex-typed behavior between age 2.5 and age 5 showed undifferentiated sex-typed behavior at age 8.

A step-down regression was then carried out to determine whether the PSAI trajectory added any

Table 4

Means and Standard Deviations of Preschool Activities Inventory Trajectory scores and Children's Activities Inventory Scores at 5-Point Intervals of Preschool Activities Inventory Score at Age 3.5 for Boys

Preschool Activities Inventory score (3.5 years)		Preschool Activities Inventory trajectory		Children's Activities Inventory (8 years)	
Category	*n*	*M*	*SD*	*M*	*SD*
≤ 44	69	1.38	1.89	56.35	10.29
45–50	152	1.66	1.31	56.24	11.55
50–55	394	1.64	1.56	58.04	11.92
55–60	486	1.78	1.47	58.48	11.92
60–65	628	2.08	1.47	60.75	10.73
65–70	443	2.18	1.59	60.90	11.31
70–75	357	2.10	1.66	60.92	11.62
75–80	115	2.55	1.59	62.23	11.02
> 80	61	2.82	1.61	62.79	11.48
Total	2,705	1.98	1.56	59.70	11.51

Table 5

Means and Standard Deviations of Preschool Activities Inventory Trajectory Scores and Children's Activities Inventory Scores at 5-Point Intervals of Preschool Activities Inventory Score at Age 3.5 for Girls

Preschool Activities Inventory score (3.5 years)		Preschool Activities Inventory trajectory		Children's Activities Inventory (8 years)	
Category	*N*	*M*	*SD*	*M*	*SD*
≤ 44	86	−4.22	1.53	37.58	12.83
45–50	142	−3.84	1.33	37.17	11.98
50–55	394	−3.41	1.35	37.59	12.10
55–60	462	−3.38	1.43	38.23	11.51
60–65	608	−2.81	1.41	39.67	12.10
65–70	568	−2.60	1.42	40.09	12.16
70–75	264	−2.20	1.43	43.44	12.40
75–80	155	−1.87	1.75	44.36	12.41
> 80	59	−0.82	1.87	51.21	13.93
Total	2,738	−2.89	1.58	39.90	12.39

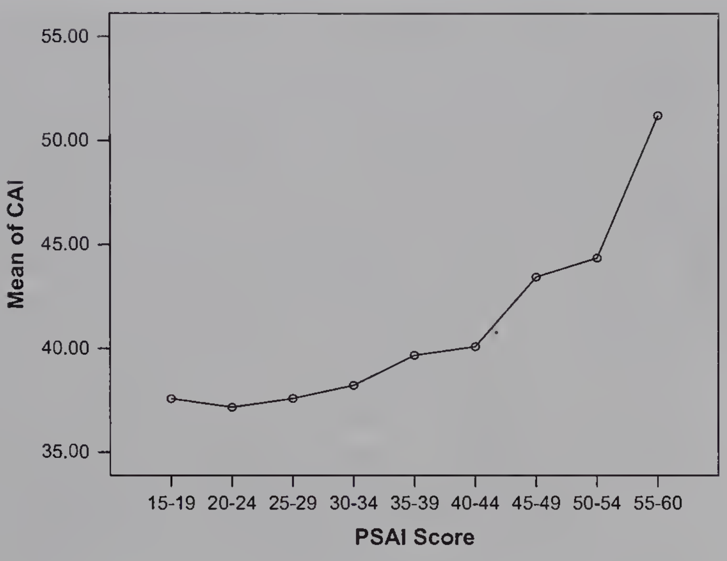


Figure 4. Mean of Children's Activities Inventory (CAI) scores plotted against Preschool Activities Inventory (PSAI) scores at age 3.5 for girls.

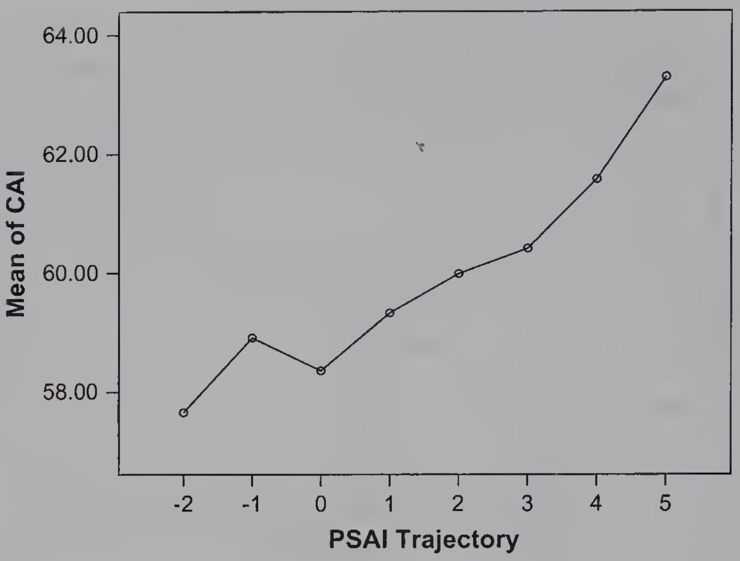


Figure 5. Mean of Children's Activities Inventory (CAI) scores plotted against Preschool Activities Inventory (PSAI) trajectory scores for boys.

predictive power to the PSAI score at age 3.5 in estimating CAI scores at age 8 (again separately for boys and girls). For boys, the multiple correlation showed a statistically significant increase from .134 to .152 ($p < .001$). For girls, the multiple correlation showed a statistically significant increase from .199 to .236 ($p < .001$). These findings show that the combination of the PSAI score at age 3.5 and the PSAI trajectory from 2.5 to 5 years was a better predictor of CAI score at age 8 for both boys and girls than either measure alone.

Discussion

The findings of this longitudinal study of a general population sample not only show that children's sex-typed behavior increases through the preschool years but also that those children who were the most sex typed at age 2.5 are still the most sex typed at age 5, with girls showing a greater increase in sex-typed behavior over time than boys. In addition, those children who showed the highest levels of sex-typed behavior during the preschool years continued to do so at age 8, as assessed by a modified self-report version of the PSAI. Thus, it appears that sex-typed toddlers grow up to be sex-typed children, that is, that sex-typed behavior remains stable over time, at least from ages 2.5 to 8 years.

Instead of the usual question "What causes children of the opposite sex to differ in their behavior?" the findings of the present study address the rather different question "What causes children of the same sex to show variation in sex-typed behavior that is

Table 6

Means and Standard Deviations of Children's Activities Inventory Scores at 1-Point Intervals of Preschool Activities Inventory Trajectory Score for Boys

Preschool Activities Inventory trajectory category	*N*	*M*	*SD*
< −1	66	57.66	12.36
−1 to 0	210	58.91	12.20
0–1	417	58.34	12.33
1–2	710	59.38	11.44
2–3	640	59.94	11.13
3–4	395	60.51	10.92
4–5	193	61.55	11.13
> 5	74	63.04	10.52
Total	2,705	59.70	11.51

Table 7

Means and Standard Deviations of Children's Activities Inventory Scores at 1-Point Intervals of Preschool Activities Inventory Trajectory Score for Girls

Preschool Activities Inventory trajectory category	*N*	*M*	*SD*
< −6	67	37.00	11.16
−6 to −5	161	36.28	10.62
−5 to −4	426	38.43	11.78
−4 to −3	636	37.99	11.61
−3 to −2	702	40.27	11.97
−2 to −1	458	41.38	12.45
−1 to 0	197	44.16	14.16
> 0	91	49.03	15.02
Total	2,738	39.90	12.39

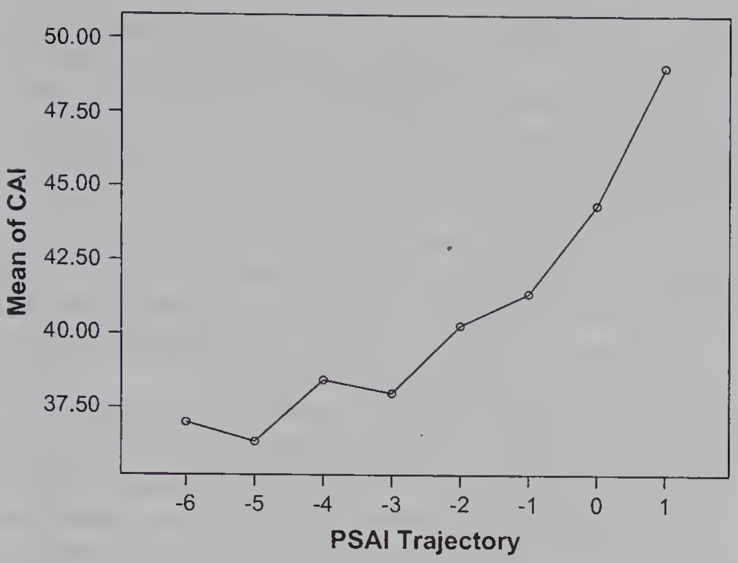


Figure 6. Mean of Children's Activities Inventory (CAI) scores plotted against Preschool Activities Inventory (PSAI) trajectory scores for girls.

apparent from an early age, intensifies over time, and remains stable from toddlerhood to middle childhood?" From a biological perspective, it could be argued that variation in prenatal androgen levels predispose some children within each sex to be more sex typed than others. The role of androgens in gender development has been studied in children exposed to atypical levels prenatally, with higher levels of testosterone associated with more male-typical toy, playmate, and activity preferences, as well as temperamental characteristics (Berenbaum & Hines, 1992; Dittmann et al., 1990; Hines & Kaufman, 1994; Pasterski et al., 2007). In a study of children in the current sample at age 3.5, an association was found between maternal testosterone concentrations during pregnancy and degree of sex-typed behavior in girls (Hines et al., 2002). Because prenatal androgen exposure is thought to influence subsequent sex-typed behavior by permanent alterations in brain development (Arnold & Gorski, 1984; De Vries & Simerly, 2002), individual variability in prenatal testosterone exposure could cause permanent neural alterations that contribute to continuity in sex-typical behavior over time. Although this explanation could account for the onset of behavioral preferences and their maintenance, it would not explain the increase we observed in sex-typed behavior with age.

Social learning and cognitive developmental explanations of behavioral differences between the sexes may also be relevant to understanding the mechanisms involved in variation in sex-typed behavior within each sex and to the increase in sex-typed behavior over time. The differential reinforcement of sex-typed behavior in boys and girls, and the extent to which children are exposed to sex-typed models, may both be implicated, such that children who experience high levels of reinforcement for sex-typed behavior, and whose role models are highly sex typed, may exhibit higher levels of sex-typed behavior themselves. Similarly, individual differences in gender schematization, that is, in the selective processing of gender-related information (Levy & Carter, 1989), and in gender labeling (Martin, Ruble, & Szkrybalo, 2002), may also play a part in within-sex variation in sex-typed behavior. Insofar as the mechanisms described by social learning and cognitive developmental theorists continue through childhood, it might be expected that their influence on children's sex-typed behavior would increase over time.

The prediction of a greater increase in sex-typed behavior by those who were most extreme to start with was also confirmed. This is of particular interest as the phenomenon of regression to the mean would lead to the opposite prediction, that is, to the prediction that those who were most extreme to start with would show the greatest decrease in sex-typed behavior over time. An explanation for our finding is that the acquisition of sex-typed behavior involves a series of interacting processes such that children who are biologically predisposed to show higher levels of sex-typed behavior as toddlers experience greater exposure to social and psychological influences that are likely to increase sex-typed behavior even further as they grow up. For example, gender segregation and the differences in play styles experienced by boys and girls in same-sex peer groups (Maccoby, 1998) may be involved, such that children who show higher levels of sex-typed behavior at the outset may be more likely to play with children of the same sex and thus be more likely to engage in sex-typed behavior. As a result, sex-typed toddlers may become even more sex-typed through childhood. Our finding that the PSAI score at age 3.5 in combination with the PSAI trajectory from 2.5 to 5 years was a better predictor of CAI score at age 8 than either measure alone supports this perspective. Further evidence for this explanation comes from the findings of Martin and Fabes (2001), who demonstrated that playing with same-sex peers directly increased sex-typed behavior over and above individual tendencies to do so. It is not clear why highly sex-typed girls show a greater increase in sex-typed behavior as they grow older than do highly sex-typed boys. However, Martin and Fabes have argued that girls' groups experience greater socialization pressures from adults than do boys' groups, which may account for the greater increase in the sex-typed behavior of highly sex-typed girls than in the sex-typed behavior of highly sex-typed boys found in the present study.

The present study is the first to follow up a large number of children from toddlerhood to middle childhood to examine the development of sex-typed behavior over time. Particular advantages of the investigation were the involvement of a general population sample of boys and girls at four time points ranging from 2.5 to 8 years and the use of an assessment instrument specifically designed to assess within-sex variation in sex-typed behavior. The findings demonstrate that the degree of sex-typed behavior exhibited by boys and girls early in life is a good indicator of their degree of sex-typed behavior as they grow up, that sex-typed behavior increases between toddlerhood and middle childhood for both boys and girls, and that divergence between sex-typical and sex-atypical children increases as they progress through childhood, particularly among girls. Our findings argue for the importance of social-cognitive processes in elaborating early sex-typed behavior, perhaps building on early tendencies set in place by the prenatal hormonal environment.

References

Arnold, A. P., & Gorski, R. A. (1984). Gonadal steroid induction of structural sex differences in the central nervous system. *Annual Review of Neuroscience*, *7*, 413–442.

Baker, D., Morris, S., & Taylor, H. (1997). *A census comparison to assess the representativeness of the ALSPAC sample*. Unpublished manuscript, University of Bristol, England.

Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice Hall.

Beere, C. (1990). *Gender roles: A handbook of tests and measures*. New York: Greenwood Press.

Berenbaum, S. A., & Hines, M. (1992). Early androgens are related to childhood sex-typed toy preferences. *Psychological Science*, *3*, 203–206.

Bussey, K., & Bandura, A. (1999). Social cognitive theory of gender development and differentiation. *Psychological Review*, *106*, 676–713.

Crick, N., Casas, J., & Mosher, M. (1997). Relational and overt aggression in preschool. *Developmental Psychology*, *33*, 579–588.

De Vries, G. J., & Simerly, R. B. (2002). Anatomy, development, and function of sexually dimorphic neural circuits in the mammalian brain. In D. W. Pfaff, A. P. Arnold, A. M. Etgen, S. E. Fahrbach, & R. T. Rubin (Eds.), *Hormones, brain and behavior* (pp. 137–191). San Diego, CA: Academic Press.

Dittmann, R. W., Kappes, M. H., Kappes, M. E., Börger, D., Stegner, H., Willig, R. H., et al. (1990). Congenital Adrenal Hyperplasia I: Gender-related behavior and attitudes in female patients and sisters. *Psychoneuroendocrinology*, *15*, 401–420.

Dunn, J., & Hughes, C. (2001). "I got some swords and you're dead!" Violent fantasy, antisocial behavior, friendship, and moral sensibility in young children. *Child Development*, *72*, 491–505.

Fabes, R., Martin, C., & Hanish, L. (2003). Young children's play qualities in same-, other-, and mixed-sex peer groups. *Child Development*, *74*, 921–932.

Golding, J., Pembrey, M., Jones, R., & the ALSPAC Study Team. (2001). ALSPAC The Avon Longitudinal Study of Parents and Children: I. Study methodology. *Paediatric and Perinatal Epidemiology*, *15*, 74–87.

Golombok, S., & Fivush, R. (1994). *Gender development*. New York: Cambridge University Press.

Golombok, S., & Rust, J. (1993a). The measurement of gender role behavior in pre-school behavior. A research note. *Journal of Child Psychology and Psychiatry*, *34*, 805–811.

Golombok, S., & Rust, J. (1993b). The Pre-School Activities Inventory: A standardised assessment of gender role in children. *Psychological Assessment*, *5*(2), 131–136.

Golombok, S., & Rust, J. (2007). *Manual of the Pre-School Activities Inventory*. London: Harcourt.

Hines, M. (2004). *Brain gender*. Oxford, UK: Oxford University Press.

Hines, M., Golombok, S., Rust, J., Johnston, K., Golding, J., & the ALSPAC Study Team (2002). Testosterone during pregnancy and gender role behaviour of preschool children: A longitudinal population study. *Child Development*, *73*, 1678–1687.

Hines, M., & Kaufman, F. R. (1994). Androgen and the development of human sex-typical behavior: Rough-and-tumble play and sex of preferred playmates in children with congenital adrenal hyperplasia (CAH). *Child Development*, *65*, 1042–1053.

Howes, C. (1988). Peer interaction among children. *Monographs of the Society for Research in Child Development*, *53* (Serial No. 217), 1–92.

Kohlberg, L. (1966). A cognitive-developmental analysis of children's sex-role concepts and attitudes. In E. E. Maccoby (Ed.), *The development of sex differences* (pp. 82–173). Stanford, CA: Stanford University Press.

LaFreniere, P., Strayor, F., & Gauthier, R. (1984). The emergence of same-sex affiliative preference among preschool peers: A developmental/ethological perspective. *Child Development*, *55*, 1958–1965.

Lever, J. (1976). Sex differences in the games children play. *Social Problems*, *23*, 478–487.

Levy, G. D., & Carter, D. B. (1989). Gender schema, gender constancy and gender role knowledge: The roles of cognitive factors in preschoolers' gender-role stereotypic attitudes. *Developmental Psychology*, *25*, 444–449.

Maccoby, E. E. (1998). *The two sexes: Growing up apart, coming together*. Cambridge, MA: Harvard University Press.

Maccoby, E. E. (2002). Gender and group process: A developmental perspective. *Current Directions in Psychological Sciences*, *11*, 54–58.

Maccoby, E. E., & Jacklin, C. N. (1987). Gender segregation in children. In H. W. Reece (Ed.), *Advances in child* *development and behavior* (pp. 239–287). New York: Academic Press.

Martin, C. L. (1993). New directions for assessing children's gender knowledge. *Developmental Review, 13*, 184–202.

Martin, C. L., & Fabes, R. A. (2001). The stability and consequences of young children's same-sex peer interactions. *Developmental Psychology, 37*, 431–446.

Martin, C. L., Ruble, D. N., & Szkrybalo, J. (2002). Cognitive theories of early gender development. *Psychological Bulletin, 128*, 903–933.

Martin, C. L., Wood, C. H., & Little, J. K. (1990). The development of gender stereotype components. *Child Development, 61*, 1891–1904.

Martin, C. L., Wood, C. H., & Little, J. K. (1990). The development of gender stereotype components. *Child Development, 61*, 1891–1904.

McNeilly-Choque, M., Hart, C., Robinson, C., Nelson, L., & Olsen, S. (1996). Overt and relational aggression on the playground: Correspondence among different informants. *Journal of Research in Childhood Education, 11*, 47–67.

Mischel, W. (1966). A social learning view of sex differences in behaviour. In E. E. Maccoby (Ed.), *The development of sex differences* (pp. 57–81). Stanford, CA: Stanford University Press.

Mischel, W. (1970). Sex-typing and socialization. In P. Mussen (Ed.), *Carmichael's manual of child psychology* (pp. 3–72). New York: Wiley.

O'Brien, M., & Huston, A. C. (1985). Activity level and sex-stereotyped toy choice in toddler boys and girls. *Journal of Genetic Psychology, 146*, 527–533.

Pasterski, V. L., Hindmarsh, P., Geffner, M., Brook, C. D. G., Brain, C., & Hines, M. (2007). Increased aggression and activity level in 3- to 11-year-old girls with congenital adrenal hyperplasia. *Hormones and Behavior, 76*, 264–278.

Pitcher, E., & Shultz, L. (1983). *Boys and girls at play: The development of sex roles*. South Hadley, MA: Bergin & Garvey.

Ruble, D. N., Martin, C. L., & Berenbaum, S. A. (2006). Gender development. In N. Eisenberg (Ed.), *Handbook of child psychology* (Vol. 3, pp. 858–932). New York: Wiley.

Serbin, L., Powlishta, K., & Gulko, J. (1993). The development of sex-typing in middle childhood. *Monographs of the Society for Research in Child Development, 58* (Serial No. 232), 955–963.

Servin, A., Bohlin, G., & Berlin, L. (1999). Sex differences in 1-3-, and 5-year-olds' toy-choice in a structured play-session. *Scandinavian Journal of Psychology, 40*, 43–48.

Sheldon, A. (1990). Pickle fights: Gendered talk in preschool disputes. *Discourse Processes, 13*, 5–31.

Snow, M. E., Jacklin, C. N., & Maccoby, E. E. (1983). Sex-of-child differences in father-child interaction at one year of age. *Child Development, 49*, 227–232.

Sutton-Smith, B., & Rosenberg, B. (1971). Sixty years of historical change in the game preferences of American children. In R. E. Herron & D. Sutton-Smith (Eds.), *Child's play.* New York: Wiley.

Whiting, B., & Edwards, C. (1988). *Children of different worlds: The formation of social behavior.* Cambridge, MA: Harvard University Press.

Manuscripts Accepted for Publication

In Defense of Qualitative Changes in Development, Jerome Kagan.

Rediscovering Development in Infancy, Joseph Campos, David Witherington, David I. Anderson, Carl I. Frankel, Ichiro Uchiyama, and Marianne Barbu-Roth.

In Defense of Core Competencies, Quantitative Change, and Continuity, Paul Quinn.

Defending Qualitative Change: The View From Dynamical Systems Theory, John Spencer and Sammy Perone.

In Defense of Change Processes, Karen Adolph and Scott R. Robinson.

Methodological and Epistemological Issues in the Interpretation of Infant Cognitive Development, Ulrich Müller and Gerald Giesbrecht.

Longitudinal Effects of Student Mobility on Three Dimensions of Elementary School Engagement, Diana Gruman, Tracy Harachi, Robert Abbott, Richard Catalano, and Charles Fleming.

Neonatal Stepping in Relation to Terrestrial Optic Flow, Marianne Barbu-Roth, David Anderson, Adeline Despres, Joelle Provasi, and Joseph Campos.

A Cognitive Approach to the Development of Early Language, Susan A. Rose, Judith F. Feldman, and Jeffery J. Jankowski.

Neural Processing of Eye Gaze and Threat-Related Emotional Facial Expressions in Infancy, Stefanie Hoehl and Tricia Striano.

Fetal Substance Exposure and Cumulative Environmental Risk in an African American Cohort, Chie Yumoto, Sandra Jacobson, and Joseph Jacobson.

Postural and Object-Oriented Experiences Advance Early Reaching, Object Exploration, and Means-End Behavior, Michele A. Lobo and James C. Galloway.

Emergence of Global Shape Processing Continues Through Adolescence, K. Suzanne Scherf, Marlene Behrmann, Ruth Kimchi, and Beatriz Luna.

Automatic Activation of Adolescents' Peer-Relational Schemas: Evidence From Priming With Facial Identity, Lauri Nummenmaa, Kätlin Peets, and Christina Salmivalli.

Effects of Aggresssion on Achievement: Does Conflict With the Teacher Make It Worse? Deborah Stipek and Sarah Miles.

Children's Cortisol and the Quality of Teacher-Child Relationships in Child Care, Jared A. Lisonbee, Jacquelyn Mize, Amie Lapp Payne, and Douglas A. Granger.

Face Adaptation and Attractiveness Aftereffects in 8-Year-Olds and Adults, Gizelle Anzures, Catherine Mondloch, and Christine Lackner.

Time Course of Visual Attention in Infant Categorization of Cats Versus Dogs: Evidence for a Head Bias as Revealed Through Eye Tracking, Paul C. Quinn, Matthew M. Doran, Jason E. Reiss, and James E. Hoffman.

Early Histories of School-Aged Children With Attention Deficit/Hyperactivity Disorder, Irene M. Loe, Maria D. Balestrino, Randall A. Phelps, Marcia Kurs-Lasky, Diego Chaves-Gnecco, Jack L. Paradise, and Heidi M. Feldman.

Maternal Depression, Children's Attachment Security, and Representational Development: An Organizational Perspective, Sheree Toth, Dante Cicchetti, Fred A. Rogosch, and Melissa Sturge-Apple.

Group Nous and Social Exclusion: The Role of Theory of Social Mind, Multiple Classification Skill, and Experience of Peer Relations Within Groups, Dominic Abrams, Adam Rutland, Joseph Pelletier, and Jennifer M. Ferrell.

Age Differences in Future Orientation and Delay Discounting, Laurence Steinberg, Sandra Graham, Lia O'Brien, Jennifer Woolard, Elizabeth Cauffman, and Marie Banich.

Children's Responses to Interparental Conflict: A Meta-Analysis of Their Associations With Child Adjustment, Kimberly A. Rhoades.

Romantic Partner Selection and Socialization During Early Adolescence, Valerie A. Simon, Julie Wargo Aikins, and Mitchell J. Prinstein.

The Social Ecology of Adolescent Alcohol Misuse, Susan Ennett, Vangie A. Foshee, Karl E. Bauman, Andrea Hussong, Robert Faris, John Hipp, L i Cai, Robert DuRant, and Heathe Luz McNaughton Reyes.

Adrenocortical Underpinnings of Children's Psychological Reactivity to Interparental Conflict, Patrick Davies, Melissa Sturge-Apple, Dante Cicchetti, and E. Mark Cummings.

Stress Regulation in Adolescents: Physiological Reactivity During the Adult Attachment Interview and Conflict Interaction, Mariëlle D. Beijersbergen, Marian J. Bakermans-Kranenburg, Marinus H. van IJzendoorn, and Femmie Juffer.

Promoting Academic and Social-Emotional School Readiness: The Head Start REDI Program, Karen L. Bierman, Celene E. Domitrovich, Robert L. Nix, Scott D. Gest, Janet A. Welsh, Mark T. Greenberg, Clancy B. Blair, Keith E. Nelson, and Suhkdeep Gill.

Mother–Infant Vagal Regulation in the Face-to-Face Still-Face Paradigm in Moderated by Maternal Sensitivity, Ginger A. Moore, Ashley L. Hill-Soderlund, Cathi B. Propper, Susan D. Calkins, W. Roger Mills-Koonce, and Martha J. Cox.

How Adolescents Come to See Themselves as More Responsible Through Participation in Youth Programs, Dustin Wood, Reed Larson, and Jane Brown.

Children's Trust in Previously Inaccurate Informants Who Were Well or Poorly Informed: When Past Errors Can Be Excused, Erika Nurmsoo and Elizabeth Robinson.

Trumping Shame by Blasts of Noise: Narcissism, Self-Esteem, Shame, and Aggression in Young Adolescents, Sander Thomaes, Brad J. Bushman, Hedy Stegge, and Tjeert Olthof.

Childhood Maltreatment and the Development of Relational and Physical Aggression: The Importance of a Gender-Informed Approach, Crystal Cullerton-Sen, Adam R. Cassidy, Dianna Murray-Close, Dante Cicchetti, Nicki R. Crick, and Fred A. Rogosch.

Temperament, Harsh and Indulgent Parenting, and Chinese Children's Proactive and Reactive Aggression, Yiyuan Xu, Jo Ann M. Farver, and Zengxiu Zhang.

A Horse of a Different Color: Specifying With Precision Infants' Mappings of Novel Nouns and Adjectives, Amy Booth and Sandra R. Waxman.

The Development of Children's Ability to Use Evidence to Infer Reality Status, Ansley Tullos and Jacqueline D. Woolley.

Changes in Stress Perception and Coping During Adolescence: The Role of Situational and Personal Factors, Inge Seiffe-Krenke, Kaisa Aunola, and Jari-Erik Nurmi.

A Process Model of Attachment-Friend Linkages: Hostile Attribution Biases and Language Ability as Intervening Mechanisms, Nancy McElwain, Cathryn Booth-LaForce, Jennifer Lansford, Xiaoying Wu, and W. Justin Dyer.

Adolescents' and Parents' Evaluations of Helping Versus Fulfilling Personal Desires in Family Situations, Judith Smetana, Marina Tasopoulos-Chan, Denise C. Gettman, Myriam Villalobos, Nicole Campione-Barr, and Aaron Metzger.

Individualizing Student Instruction Precisely: Effects of Child by Instruction Interactions on First Graders' Literacy Development, Carol Connor, Shayne B. Piasta, Stephanie Glasney, Barry Fishman, Christopher Schatschneider, Phyllis Underwood, Elizabeth Coyne Crowe, and Frederick J. Morrison.

Developmental Psychology Lecturer

Department of Psychology, Auckland
Vacancy Number A173-08L

The Department of Psychology invites applications for a tenure-track Lecturer/Senior Lecturer in Developmental Psychology. Applicants should have a PhD in Psychology and the qualifications necessary to teach Developmental Psychology at both undergraduate and postgraduate levels. You should also have a proven record of research and be able to supervise post graduate students. The Department is one of the largest in New Zealand, with 920 equivalent full-time students, 40 permanent academic staff and 10 administrative and technical support staff. The Department offers Bachelor, Honours, Masters and PhD degrees in Psychology and postgraduate qualifications in professional psychology.

Ideally, the role will start 1 February 2009.

Applications close Friday, 31 October 2008.

For further information contact Associate Professor Fred Seymour, Head of Department of Psychology on +64 9 373 7599 ext 88414, or email f.seymour@auckland.ac.nz.

The department's website is www.psych.auckland.ac.nz.

Please apply online at www.opportunities.auckland.ac.nz.

THE UNIVERSITY
OF AUCKLAND
NEW ZEALAND
Te Whare Wānanga o Tāmaki Makaurau

All Wiley-Blackwell journals are now available through Wiley InterScience!

WILEY
InterScience®

Visit **www.interscience.wiley.com**
for one-stop access to more than 3 million articles across 1400 titles, as well as over 6000 online books, plus databases, reference works, protocols and other electronic resources.

WILEY-
BLACKWELL

www.interscience.wiley.com

Notice to Contributors

Child Development publishes empirical, theoretical, review, applied, and policy articles reporting research on child development. Published by the interdisciplinary Society for Research in Child Development (SRCD), the journal welcomes relevant submissions from all disciplines. Further information is available at http://www.srcd.org/cd.html.

Types of Articles

Child Development considers manuscripts in formats described below. Inquiries concerning alternative formats should be addressed to the Editor prior to submission. All submissions are expected to be no more than 40 manuscript pages, including tables, references, and figures (but excluding appendices). Authors should provide a justification if the submission is substantially longer. Unless the editor finds that justification compelling, the submission will be returned to the author for shortening prior to editorial review.

Empirical articles comprise the major portion of the journal. To be accepted, empirical articles must be judged as being high in scientific quality, contributing to the empirical base of child development, and having important theoretical, practical, or interdisciplinary implications. Reports of multiple studies, methods, or settings are encouraged, but single-study reports are also considered. Empirical articles will thus vary considerably in length, but should be no longer than 40 manuscript pages; text and graphics should be as concise as material permits. All modes of empirical research are welcome.

Empirical reports are reserved for short cutting-edge empirical papers that are no longer than 4,000 words in length (including text, tables, footnotes, appendices), which advance research and knowledge in an area through noteworthy findings and/or new methods. For manuscripts that require longer descriptions of methods and results, authors should use the Empirical article format.

Reviews focus on past empirical and/or on conceptual and theoretical work. They are expected to synthesize or evaluate a topic or issue relevant to child development, should appeal to a broad audience, and may be followed by a small number of solicited commentaries.

Essays describe original concepts, methods, trends, applications, and theories; these may also be accompanied by solicited commentaries.

Child Development and ... are articles that provide readers with tutorials about some new concept or academic specialty pertinent to research in child development. These papers should review the major definitions, methods, and findings of the concept or specialty and discuss past or potential links to child development.

From another perspective is a format in which papers on a focal topic, written by different authors, are published simultaneously. Papers represent diverse perspectives (e.g., authors whose work represents different populations; different disciplines; different theories, methods, or analytic tools). In some cases, calls for submissions on particular topics will be disseminated through SRCD (via e-mail or SRCD publications), and submissions will undergo normal editorial review. In some cases, a submitted manuscript (e.g., an empirical article) may be selected as a lead article for this format, with invited commentaries providing additional perspectives. The editors also welcome suggestions from readers for topics for this format.

Manuscript Submission

Child Development invites for consideration manuscripts that are neither identical to nor substantially similar to work published or under review elsewhere. In the submission cover letter, please provide details about other published or submitted papers having substantial overlap (including data sets) with the new CD submission to enable editors to judge whether the new submission is sufficiently distinct from other work to warrant consideration. Please note if the paper is posted on a website, see http://www.srcd.org/webposting.html. Editors retain the right to reject manuscripts that do not meet established ethical standards for research or dissemination.

Manuscripts should be submitted electronically to the Child Development Online Submission Site at www.srcd.org/CDsubmit/ as a Word or WordPerfect file. Please also submit a cover letter that contains the name(s) of the author(s) and affiliation(s), and the street address, telephone, fax, and electronic mail address of the corresponding author. A corresponding author's submission to *Child Development* implies that all co-authors have agreed to the content and form of the manuscript and that the ethical standards of SRCD have been followed (see the *Child Development* website or pp. 283–284 of the 2000 SRCD Directory). Any financial interest or conflict of interest must be explained to the Editor in the cover letter. The corresponding author is responsible for informing all coauthors, in a timely manner, of manuscript submission, editorial decisions, reviews, and revisions. Please note that if your submission (or its revision) is accepted for publication, each author must submit a signed Exclusive License Form (http://www.srcd.org/Documents/Publications/CD/cdcopyright.pdf) before the manuscript can enter production. Planning ahead to meet this requirement is especially important for manuscripts with multiple authors.

The manuscript file should be formatted with double-spaced, 12-point type, and should include a single paragraph abstract of 100–120 words. Please follow all guidelines on format, style, and ethics provided in the Publication Manual (5th ed.) of the American Psychological Association. Figures included with initial submissions will not be returned. Therefore, please submit only electronic files or copies of figures. Authors should keep a copy of all correspondence, files, and figures to guard against loss.

Manuscript Review

If you have any questions about your submission, please inquire at cdev@srcd.org or call (734) 926-0615. Each manuscript is handled by the Editor or an Associate Editor who consults with one or more Consulting Editors and/or ad hoc reviewers who have relevant expertise. To ensure blind review, cover sheets are removed before review; authors should avoid including any other information about identity or affiliation in submissions. Copies of the submission and associated correspondence are retained in the SRCD archives. For accepted manuscripts, authors are required to prepare a 300–500 layperson's summary for public dissemination purposes. Details are provided to authors as part of final processing.

There is no charge for publication in *Child Development* unless tabular or graphic materials exceed 10% of the total number of pages. Charges are also levied for changes in proofs other than correction of printer's errors. Any inquiries relating to charges or business matters (including reprint orders) should be addressed to Wiley-Blackwell, *Child Development*, Production Coordinator, 350 Main Street, Malden, MA 02148, (781) 388-8200.

Inquiries and suggestions regarding editorial policy may be addressed to: Dr. Jeffrey J. Lockman, Incoming Editor, Department of Psychology, Tulane University, 2007 Percival Stern Hall, New Orleans, LA 70118-5698, lockman@tulane.edu. *Please do not send electronic or paper manuscripts to the above address.* Doing so will delay the processing of your manuscript.

Monographs of the Society for Research in Child Development

Recent

Understanding Mother-Adolescent Conflict Discussions: Concurrent and Across-Time Prediction From Youths' Dispositions and Parenting—*Nancy Eisenberg, Claire Hofer, Tracy L. Spinrad, Elizabeth T. Gershoff, Carlos Valiente, Sandra Losoya, Qing Zhou, Amanda Cumberland, Jeffrey Liew, Mark Reiser, and Elizabeth Maxon* (Serial No. 290, 2008)

Developing Object Concepts in Infancy: An Associative Learning Perspective—*David H. Rakison and Gary Lupyan* (Serial No. 289, 2008).

The Genetic and Environmental Origins of Learning Abilities and Disabilities in the Early School Years—*Yulia Kovas, Claire M. A. Haworth, Philip S. Dale, and Robert Plomin* (Serial No. 288, 2007).

The Preservation of Two Infant Temperaments Into Adolescence—*Jerome Kagan, Nancy Snidman, Vali Kahn, and Sara Towsley* (Serial No. 287, 2007).

Children's Questions: A Mechanism for Cognitive Development—*Michelle M. Chouinard* (Serial No. 286, 2007).

Best Practices in Quantitative Methods for Developmentalists—*Kathleen McCartney, Margaret R. Burchinal, and Kristen L. Bub* (Serial No. 285, 2006).

Foundations of Self-Awareness: An Exploration Through Autism—*R. Peter Hobson, Gayathri Chidambi, Anthony Lee, and Jessica Meyer* (Serial No. 284, 2006).

Our Children Too: A History of the Black Caucus of the Society for Research in Child Development, 1973–1997—*Diana T. Slaughter-Defoe, Aline M. Garrett, and Algea O. Harrison-Hale* (Serial No. 283, 2006).

Parental Support, Psychological Control, and Behavioral Control: Assessing Relevance Across Time, Culture, and Method—*Brian K. Barber, Heidi E. Stolz, and Joseph A. Olsen* (Serial No. 282, 2005).

Being Hurt and Hurting Others: Children's Narrative Accounts and Moral Judgments of Their Own Interpersonal Conflicts—*Cecilia Wainryb, Beverly A. Brenl, and Sonia Matwin* (Serial No. 281, 2005).

Childhood Sexual Assault Victims: Long-Term Outcomes After Testifying in Criminal Court—*Jodi A. Quas, Gail S. Goodman, Simona Ghetti, Kristen Weede Alexander, Robin Edelstein, Allison D. Redlich, Ingrid M. Cordon, and David P. H. Jones* (Serial No. 280, 2005).

The Emergence of Social Cognition in Three Young Chimpanzees—*Michael Tomasello and Malinda Carpenter* (Serial No. 279, 2005).

Trajectories of Physical Aggression From Toddlerhood to Middle Childhood: Predictors, Correlates, and Outcomes—*NICHD Early Child Care Research Network* (Serial No. 278, 2004).

Constraints on Conceptual Development: A Case Study of the Acquisition of Folkbiological and Folksociological Knowledge in Madagascar—*Rita Astuti, Gregg E. A. Solomon, and Susan Carey* (Serial No. 277, 2004).

Origins and Early Development of Human Body Knowledge—*Virginia Slaughter, Michelle Heron, Linda Jenkins, and Elizabeth Tilse* (Serial No. 276, 2004).

Mother – Child Conversations About Gender: Understanding the Acquisition of Essentialist Beliefs—*Susan A. Gelman, Marianne G. Taylor, and Simone P. Nguyen* (Serial No. 275, 2004).

The Development of Executive Function in Early Childhood—*Philip David Zelazo, Ulrich Müller, Douglas Frye, and Stuart Marcovitch* (Serial No. 274, 2003).

Personal Persistence, Identity Development, and Suicide: A Study of Native and Non-native North American Adolescents—*Michael J. Chandler, Christopher E. Lalonde, Bryan W. Sokol, and Darcy Hallett* (Serial No. 273, 2003).

Personality and Development in Childhood: A Person-Centered Approach—*Daniel Hart, Robert Atkins, and Suzanne Fegley* (Serial No. 272, 2003).

How Children and Adolescents Evaluate Gender and Racial Exclusion—*Melanie Killen, Jennie Lee-Kim, Heidi McGlothlin, and Charles Stangor* (Serial No. 271, 2002).

Child Emotional Security and Interparental Conflict—*Patrick T. Davies, Gordon T. Harold, Marcie C. Goeke-Morey, and E. Mark Cummings* (Serial No. 270, 2002).

The Developmental Course of Gender Differentiation: Conceptualizing, Measuring, and Evaluating Constructs and Pathways—*Lynn S. Liben and Rebecca S. Bigler* (Serial No. 269, 2002).

The Development of Mental Processing: Efficiency, Working Memory, and Thinking—*Andreas Demetriou, Constantinos Christou, George Spanoudis, and Maria Platsidou* (Serial No. 268, 2002).

SRCD
Society for Research in Child Development

WILEY-BLACKWELL

Order from Wiley-Blackwell at subscrip@bos.blackwellpublishing.com

This journal is available online. Contact your librarian or visit www.blackwell-synergy.com.

CURRENT YR/VOL

Marygrove College Library
8425 West McNichols Road
Detroit, MI 48221

Child Development

me 79 Number 6

mber/December 2008

Editor

Lynn S. Liben

Associate Editors

Joseph P. Allen
Neil E. Berthier
Marvin W. Daehler
Nancy E. Hill
Kathy A. Hirsh-Pasek
Campbell Leaper
Patricia H. Miller
Sharon Nelson-Le Gall
Paul C. Quinn
Theresa A. Thorkildsen

Published by Wiley-Blackwell for the Society for Research in Child Development

Published by Wiley-Blackwell for the Society for Research in Child Development

Child Development

CHILD DEVELOPMENT (ISSN 0009-3920), one of three publications of the Society for Research in Child Development, Inc., is published bimonthly by Wiley Subscription Services, Inc., a Wiley Company, 111 River St., Hoboken, NJ 07030-5774. Periodical Postage Paid at Hoboken, NJ and additional offices.

Postmaster: Send all address changes to *CHILD DEVELOPMENT*, Journal Customer Services, John Wiley & Sons Inc., 350 Main Street, Malden, MA 02148-5020.

Information for Subscribers: *Child Development* is published in six issues per year. Subscription prices for 2008 are: Premium Institutional: US$585 (The Americas), £417 (Rest of World). Premium Online-Only rates are also available (see below). Prices are exclusive of tax. Australian GST, Canadian GST and European VAT will be applied at the appropriate rates. For more information on current tax rates, please go to www.wiley.com, click on Help and follow the link through to Journal subscriptions. The Premium institutional price includes online access to the current and all online back files to January 1st 1997, where available. For other pricing options, including access information and terms and conditions, please visit www.interscience.wiley.com/journals.

Delivery Terms and Legal Title: Prices include delivery of print journals to the recipient's address. Delivery terms are Delivered Duty Unpaid (DDU); the recipient is responsible for paying any import duty or taxes. Legal title passes to the customer on despatch by our distributors.

For more information about Wiley-Blackwell journals, including online access information, terms and conditions, and other pricing options, please visit www.blackwellpublishing.com or contact our customer service department, tel: (800) 835-6770 or +1 (781) 388-8598 (US office); +44 (0) 1865 778315 (UK office); or +65 6511 8000 (Asia Office). Subscription in 2008 includes three issues of *Monographs of the Society for Research in Child Development* and three issues of *Child Development Perspectives*. Single issues from current and recent volumes are available at the current single issue price from Wiley-Blackwell Journals. Earlier issues may be obtained from Periodicals Service Company, 11 Main Street, Germantown, NY 12526, USA. Tel: +1 518 537 4700, Fax: +1 518 537 5899, Email: psc@periodicals.com. Volumes 37 to date are available in microfilm from ProQuest Information and Learning, Serials Customer Service Department, 300 North Zeeb Road, Ann Arbor, MI 48106; telephone: (800) 521-0600 x2873.

Electronic Access: The *Child Development* website can be found at www.srcd.org/cd.html. Abstract information for this journal is electronically available at www.blackwellpublishing.com/cdev. This journal is available online at Wiley InterScience. Visit www3.interscience.wiley.com to search the articles and register for table of contents e-mail alerts.

For new orders, renewals, sample copy requests, claims, changes of address and all other subscription correspondences please contact the Journals Department at your nearest Wiley-Blackwell office (address details listed above). UK office phone: +44 (0) 1865-778315, Fax: +44 (0) 1865-471775, Email: cs-journals@wiley.com; US office phone: 800-835-6770 or 781-388-8598, Fax: 781-388-8232, Email: cs-journals@wiley.com; Asia office phone: +65 6511 8000, Fax: +44 (0) 1865-471775, Email: cs-journals@wiley.com.

Membership communication and address changes should be addressed to the SRCD Executive Office, 2950 South State Street, Suite 401, Ann Arbor, MI 48104, USA. http://www.srcd.org/

Copyright and Photocopying: Journal Compilation © 2008 The Society for Research in Child Development, Inc. All rights reserved. No part of this publication (apart from articles marked "OnlineOpen") may be reproduced, stored or transmitted in any form or by any means without the prior permission in writing from the copyright holder. Authorization to photocopy items for internal and personal use is granted by the copyright holder for libraries and other users registered with their local Reproduction Rights Organization (RRO), e.g., Copyright Clearance Center (CCC), 222 Rosewood Drive, Danvers, MA 01923, USA (www.copyright.com), provided the appropriate fee is paid directly to the RRO. This consent does not extend to other kinds of copying such as copying for general distribution, for advertising and promotional purposes, for creating new collective works or for resale. OnlineOpen articles are made available in accordance with the terms of the Creative Commons Deed, Attribution 2.5 (further details available from www.creativecommons.org), which allows Open Access dissemination of the article, but does not permit commercial exploitation or the creation of derivative works. Special requests should be addressed to jrights@wiley.com.

For advertising information, please visit the journal's website at www.blackwellpublishing.com/cdev or contact the Academic and Science Advertising Sales Coordinator at journaladsUSA@bos.blackwellpublishing.com. 350 Main St., Malden, MA 02148. Phone: 781.388.8532, Fax: 781.338.8532.

Indexing/Abstracting: The contents of this journal are indexed or abstracted in the following: *Academic ASAP; Expanded Academic ASAP; Academic Search Elite; Academic Search Premier; Applied Social Sciences Index and Abstracts; Biomedical Reference Collection: Comprehensive Edition; CAB Abstracts; CABDirect; Contents Pages in Education; Current Contents/Social & Behavioral Science; Current Index to Journals in Education; Cambridge Scientific Abstracts Linguistics & Language Behavior Abstracts; CatchWord; Current Index to Journals in Education; Educational Research Abstracts; Environmental Sciences & Pollution Management; e-psyche; EBSCO Sociological Collection; EBSCO Online; EBSCO Professional Development Collection; Educational Research Abstracts; Educational Resources Information Center; Health and Safety Science Abstracts; ISI Basic Social Sciences Index; InfoTrac OneFile; Ingenta; JCR Social Sciences Edition; Leisure, Recreation, and Tourism Abstracts; Online Computer Library Center FirstSearch Electronic Collections Online; Periodical Abstracts Library; Periodical Abstracts Research II; ProQuest 5000; ProQuest Social Science Plus Text; PsycINFO; Psychological Abstracts; Psychology and Behavioral Sciences Collection; Pollution Abstracts; Safety Science & Risk Abstracts; Sage Family Studies Abstracts; Social Services Abstracts; Social Science Citation Index; Sociological Abstracts; Student Resource Center Bronze; Student Resource Center Gold; Student Resource Center Silver; Violence & Abuse Abstracts; Wilson OmniFile V; Wilson Social Sciences Abstracts;* and *Wilson Social Sciences Index.*

The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials, ANSI Z39.48—1984. This journal is printed on acid-free paper.

Published by Wiley-Blackwell for the Society for Research in Child Development

Child Development

Editor

Lynn S. Liben
The Pennsylvania State University

Associate Editors

Joseph P. Allen, *University of Virginia*
Neil E. Berthier, *University of Massachusetts*
Marvin W. Daehler, *University of Massachusetts*
Nancy E. Hill, *Duke University*
Kathy A. Hirsh-Pasek, *Temple University*
Campbell Leaper, *University of California, Santa Cruz*
Patricia H. Miller, *The University of Georgia*
Sharon Nelson-Le Gall, *University of Pittsburgh*
Paul C. Quinn, *University of Delaware*
Theresa A. Thorkildsen, *University of Illinois at Chicago*

Managing Editor

Detra N. Davis

Editorial Assistant

Kaitlin MacKenzie

Consulting Editors

Brian P. Ackerman
Daisuke Akiba
Eric M. Anderman
Roger Bakeman
Lynne E. Baker-Ward
Karen Bartsch
Amy D. Bellmore
Ramesh S. Bhatt
Rebecca S. Bigler
Pamela Blewitt
Celia A. Brownell
Peter Bryant
Kristin Buss
Susan D. Calkins
Tara C. Callaghan
Joseph J. Campos
Deborah M. Capaldi
Noel A. Card
Zhe Chen
Pamela M. Cole
John Colombo
William E. Cross Jr.
Pamela E. Davis-Kean
Adele Diamond
Mary Ann Foley
Bridget Franks
Andrew J. Fuligni
Mary Gauvain
Susan L. Golbeck
H. Hill Goldsmith
Roberta M. Golinkoff
Douglas Granger
Catherine A. Haden
Laura D. Hanish
Charles C. Helwig
Gail D. Heyman
Erika Hoff
George J. Hollich
Alice Sterling Honig
Cynthia Hudley
Jennifer Jenkins
Susan Scanlon Jones
Deborah G. Kemler Nelson
Melanie Killen
Roger Kobak
Grazyna Kochanska
Brett Laursen
Bonnie Jean Leadbeater
Elizabeth Ann Lemerise
Yonata Levy
Joan M. Lucariello
Sarah C. Mangelsdorf
Denis Mareschal
Herb W. Marsh
Clay W. Mash
Susan M. McHale
Derek E. Montgomery
Ulrich Mueller
Tamera Murdock
Letitia R. Naigles
Ageliki Nicolopoulou
Thomas G. O'Connor
Yukari Okamoto
Jamie M. Ostrov
David G. Perry
Jodie M. Plumert
Gary Resnick
Lori A. Roggman
Glenn I. Roisman
Amanda J. Rose
Kenneth H. Rubin
Diane N. Ruble
Martin D. Ruck
Ted Ruffman
Dale H. Schunk
Margaret L. Signorella
Judith G. Smetana
David Sobel
Christia Spears Brown
Cynthia A. Stifter
Mike L. Stoolmiller
Susan Sy
Lisa E. Szechter
Rebecca Treiman
Georgene L. Troseth
Marion K. Underwood
Marinus H. van IJzendoorn
Brian E. Vaughn
Melissa K. Welch-Ross
Allan Wigfield
Terry Wood
Jacqueline D. Woolley
Barbara Younger-Rossmann

Child Development, November/December 2008, Volume 79, Number 6, Pages 1597–1970

Contents

EDITORIAL

Reflections on Child Development: The Journal and the Field
Lynn S. Liben 1597

ESSAYS

Introduction

Continuities and Discontinuities in Children and Scholarship
Lynn S. Liben 1600

Lead Essay

In Defense of Qualitative Changes in Development
Jerome Kagan 1606

Commentaries

Rediscovering Development in Infancy
Joseph J. Campos, David Witherington, David I. Anderson, Carl I. Frankel, Ichiro Uchiyama, and Marianne Barbu-Roth 1625

In Defense of Core Competencies, Quantitative Change, and Continuity
Paul C. Quinn 1633

Defending Qualitative Change: The View From Dynamical Systems Theory
John P. Spencer and Sammy Perone 1639

In Defense of Change Processes
Karen E. Adolph and Scott R. Robinson 1648

Methodological and Epistemological Issues in the Interpretation of Infant Cognitive Development
Ulrich Müller and Gerald Giesbrecht 1654

EMPIRICAL ARTICLES

Automatic Activation of Adolescents' Peer-Relational Schemas: Evidence From Priming With Facial Identity
Lauri Nummenmaa, Kätlin Peets, and Christina Salmivalli 1659

Romantic Partner Selection and Socialization During Early Adolescence
Valerie A. Simon, Julie Wargo Aikins, and Mitchell J. Prinstein 1676

Adrenocortical Underpinnings of Children's Psychological Reactivity to Interparental Conflict
Patrick T. Davies, Melissa L. Sturge-Apple, Dante Cicchetti, and E. Mark Cummings 1693

Stress Regulation in Adolescents: Physiological Reactivity During the Adult Attachment Interview and Conflict Interaction
Mariëlle D. Beijersbergen, Marian J. Bakermans-Kranenburg, Marinus H. van IJzendoorn, and Femmie Juffer 1707

Effects of Aggression on Achievement: Does Conflict With the Teacher Make It Worse?
Deborah Stipek and Sarah Miles 1721

Childhood Maltreatment and the Development of Relational and Physical Aggression: The Importance of a Gender-Informed Approach
Crystal Cullerton-Sen, Adam R. Cassidy, Dianna Murray-Close, Dante Cicchetti, Nicki R. Crick, and Fred A. Rogosch 1736

Neural Processing of Eye Gaze and Threat-Related Emotional Facial Expressions in Infancy
Stefanie Hoehl and Tricia Striano 1752

Fetal Substance Exposure and Cumulative Environmental Risk in an African American Cohort
Chie Yumoto, Sandra W. Jacobson, and Joseph L. Jacobson 1761

The Social Ecology of Adolescent Alcohol Misuse
Susan T. Ennett, Vangie A. Foshee, Karl E. Bauman, Andrea Hussong, Li Cai, Heathe Luz McNaughton Reyes, Robert Faris, John Hipp, and Robert DuRant 1777

Trumping Shame by Blasts of Noise: Narcissism, Self-Esteem, Shame, and Aggression in Young Adolescents
Sander Thomaes, Brad J. Bushman, Hedy Stegge, and Tjeert Olthof 1792

Promoting Academic and Social-Emotional School Readiness: The Head Start REDI Program
Karen L. Bierman, Celene E. Domitrovich, Robert L. Nix, Scott D. Gest, Janet A. Welsh, Mark T. Greenberg, Clancy Blair, Keith E. Nelson, and Sukhdeep Gill 1802

Children's Cortisol and the Quality of Teacher – Child Relationships in Child Care
Jared A. Lisonbee, Jacquelyn Mize, Amie Lapp Payne, and Douglas A. Granger 1818

Longitudinal Effects of Student Mobility on Three Dimensions of Elementary School Engagement
Diana H. Gruman, Tracy W. Harachi, Robert D. Abbott, Richard F. Catalano, and Charles B. Fleming 1833

Early Histories of School-Aged Children With Attention-Deficit/Hyperactivity Disorder
Irene M. Loe, Maria D. Balestrino, Randall A. Phelps, Marcia Kurs-Lasky, Diego Chaves-Gnecco, Jack L. Paradise, and Heidi M. Feldman 1853

Postural and Object-Oriented Experiences Advance Early Reaching, Object Exploration, and Means – End Behavior
Michele A. Lobo and James C. Galloway 1869

A Process Model of Attachment – Friend Linkages: Hostile Attribution Biases, Language Ability, and Mother – Child Affective Mutuality as Intervening Mechanisms
Nancy L. McElwain, Cathryn Booth-LaForce, Jennifer E. Lansford, Xiaoying Wu, and W. Justin Dyer 1891

Testing an Idealized Dynamic Cascade Model of the Development of Serious Violence in Adolescence
Kenneth A. Dodge, Mark T. Greenberg, Patrick S. Malone, and Conduct Problems Prevention Research Group 1907

Investigating Preschoolers' Categorical Thinking About Gender Through Imitation, Attention, and the Use of Self-Categories
Diana M. Grace, Barbara J. David, and Michelle K. Ryan 1928

Children's Responses to Interparental Conflict: A Meta-Analysis of Their Associations With Child Adjustment
Kimberly A. Rhoades 1942

OTHERS

Manuscripts Accepted 1957

Editorial Acknowledgments 1959

Index to Volume 1964

Child Development, November/December 2008, Volume 79, Number 6, Pages 1597 – 1599

Reflections on Child Development: The Journal and the Field

Lynn S. Liben
The Pennsylvania State University

One of the privileges of serving as editor of *Child Development* (*CD*) is the chance to write two editorials—one as one begins one's editorship and another as one departs.

My first editorial was filled with a sense of excitement and opportunity. My comments were focused primarily on the journal itself, what it offered, what I hoped it would continue to offer, and how I hoped it might be changed, ever so slightly, to achieve some additional goals.

Although I am perhaps too close to provide an unbiased appraisal, the journal appears to have continued on a strong trajectory. Among the signs of the journal's continued health and development have been annual increases in the numbers of article submissions (despite growing numbers of competing journals), expanded media coverage of *CD* research (traceable to the press releases prepared by the Office of Policy and Communication based on the layperson summaries now required from *CD* authors), the documented high volume of online use of articles (including many of the articles and commentaries in special collections such as the special issue on *Race, Ethnicity, and Culture in Child Development* and the *From Another Perspective* collections on the effects of day care and on emotion regulation), the initiation and refinement of streamlined and user-friendly submission and review procedures (about which we often received unsolicited praise), and data from citation indices showing *CD*'s persistent high impact in a number of fields (e.g., developmental and educational psychology).

I am more than happy to take credit for navigating the journal along this positive trajectory, but in truth, it has been a collective effort. Thus, I would like to take this opportunity to thank, again, all those who have worked with me over the past 6 years as associate editors, managing editors, production editors, consulting editors, and ad hoc reviewers, and to thank the many other staff members, Society for Research in Child Development (SRCD) officers, and SRCD committee members who have been essential to the operation of *CD*. This is not an editorship that one can do alone.

My comments so far all concern the journal itself. But the editor of *CD* also has reason to reflect on the field more broadly. The person in this role is often asked about trends in the field, about the latest discoveries, and about continuing controversies. It is on these broader issues of developmental scholarship that I have chosen to focus the remainder of my final editorial.

It is perhaps a metaphor that can best capture my view of the field and where it has traveled over time, including over the half dozen years of my editorial term. The metaphor I would select is the glass—both half full and half empty.

One way to see the optimistic version of the glass is to read with contemporary eyes the articles that appeared in *CD* during 1930, the first year of the journal's publication. This is an exercise I often assign in teaching, but I suspect that it is one even more illuminating for experienced scholars than it is for novices. The reader of the 21st century is struck by the limited scope of those early articles, by their greater reliance on what we would now characterize as unreliable and subjective observational and coding systems, by their small samples, by their unsophisticated and primarily descriptive statistics, and, not surprisingly given the relative infancy of the field, by their more restricted links to theoretical positions and other empirical findings.

In the roughly 75 years that have elapsed since those first issues of *CD* appeared, the field of child development has grown exponentially, adding content and theory, and, especially, method. Observational systems are better documented, more detailed, and reliable; technological advances allow the measurement of biological variables or events that were earlier unobservable or perhaps even unimaginable; and statistical methods and the associated computer technology to implement them have provided new ways to generate and evaluate complex models and experimental effects.

To illustrate how far the field has come methodologically, consider Myrtle McGraw's (1935) classic

---

Correspondence concerning this article should be addressed to Lynn S. Liben, Department of Psychology, 450 Moore Building, The Pennsylvania State University, University Park, PA 16802. Electronic mail may be sent to liben@psu.edu.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7906-0001

study on the effects of physical exercise on infant development. Using a cotwin design to hold genetic endowment constant, she gave Johnny—but not Jimmy—a special exercise regimen and compared the two boys' motor outcomes. It seems almost quaint to read that the categorization of the twins as monozygotic was based solely on their having shared a single placenta. As is often forgotten when the classic Johnny and Jimmy work is discussed, there were later indications that the brothers were not, in fact, monozygotic twins: "After the twins were older, study by the resemblance method cast doubt upon their single-ovum origin. . . . The differences in personal appearance have become more noticeable as the children have grown older" (McGraw, 1935, p. 33). Without question, contemporary genetic research has devised far stronger methods with which to test individuals' genetic material. It is not only measures of biological qualities that have improved. Even in the short span of my own editorship, the field has shown an increased range of topics, ever more sophisticated methods for collecting and analyzing data, increased interdisciplinarity, greater attention to diversity, and enhanced attention to the application of scholarly contributions to the lives of children as they develop in families, schools, and community.

Lest the field become too self-congratulatory, however, it is equally important to recognize that the glass remains half empty as well. Today's readers of those early issues of *CD* are also likely to be struck by the almost uncanny similarity between the questions asked then and now. Skeptical readers might find themselves wondering whether we have, in fact, made any progress at all. If one compares an article on a particular topic from an early issue to one on the same general topic in a recent issue (another teaching exercise I often employ), one is often struck by remarkably similar questions with remarkably similar and still only tentative, incomplete answers.

An important question that emerges from these historical contrasts is whether our field is as cumulative a science as it might be. As the editor of *CD*, I have seen a large number of articles describing a wide variety of research projects on a disparate range of topics using a diversity of methods. This experience has only reinforced an observation that I made roughly a decade ago (Liben, 1997) concerning the incomplete grounding of much contemporary work in past scholarship. That is, theories and findings whose core points appeared many decades ago are proposed, *de novo*, today. Contemporary investigators often appear to be unaware of an earlier version of their "new" idea, particularly when the vocabulary used to label it differs. This observation on the state of the field is neither unique to me nor is it applicable to only the recent past. For example, Cairns and Ornstein (1979) made a related point in a historical review of developmental psychology. In the course of discussing James Mark Baldwin, they said: "It is an irony in the history of developmental thought that one of its most brilliant originators is, today, virtually unknown and unrecognized. The irony lies, of course, in the presumed commitment of developmental theory to historical processes while neglecting its own scientific history" (pp. 470–471). Even earlier, Kessen (1965) commented on the field's historical neglect, saying: "The history of child study is a history of rediscovery. With remarkable regularity, the same themes appear, are elaborated for a while, then fade" (p. 1).

One means of assessing the influence of older work on contemporary scholarship is to examine the citation of earlier work in current articles. A tally of the references included in the most recent issue of *CD* available at the time this editorial was written (volume 79, issue 4) revealed that of the 1,259 references contained within 23 articles, only 4 referred to publications prior to 1960, 17 were drawn from the 1960s, and 66 from the 1970s. The remaining 1,172 reference entries were from the 1980s ($n = 223$), 1990s ($n = 437$), and 2000s ($n = 512$).

It is not, however, only that older theories and empirical work are often overlooked in contemporary literature. It is also that they are often distorted. Sometimes distortions appear to evolve as a result of using successive secondary or tertiary sources rather than consulting primary sources (for a well-documented illustration, see Harris, 1979). Sometimes distortions appear to occur as the result of erecting theoretical straw men (Liben, 1997) that—although potentially useful for clarifying hypothetical alternatives—may misrepresent constructs and the metatheoretical bases of the arguments (Overton, 2006). Whatever the reason, earlier work that is either forgotten or distorted does not provide a solid foundation on which to build incremental knowledge.

The importance of building on prior accomplishments is, of course, well appreciated in science. A phrase commonly attributed to Sir Isaac Newton (see Merton, 1985/1965) captures this notion well: "If I have seen farther, it is by standing on the shoulders of giants." I would not be the first to suggest that our field does not do this as well as it might. In one particularly dramatic statement, Zeaman (1959) compared the natural and social sciences by suggesting that "in the natural sciences, each succeeding generation stands on the shoulders of those that have gone before, while in the social sciences, each generation steps in the face of its predecessors" (p. 167). As Cairns and Ornstein (1979) noted, there is indeed irony in a developmental field that fails to build on its own developmental roots.

To leave the shoulders of giants and return to the metaphorical glass, what can be done to fill the glass of *CD* yet higher? Consistent with the evaluation offered initially with respect to the trajectories of the journal and the field, one obvious strategy is to continue to add to the corpus of high-quality work. Consistent with the argument that historical contributions are often not attended to sufficiently, a second strategy is to promote more and closer attention to earlier scholarship, even as investigators simultaneously break new ground. Consistent with the argument that earlier work is often distorted or misunderstood, a third strategy is to encourage careful readings that include attention to primary sources and to metatheoretical issues. It is not only to honor our intellectual forbearers. It is also to benefit from them and to increase the cumulative nature of developmental science.

In closing my final editorial, I would like to leave the lofty realm of pronouncements for the entire field and return to a more personal level. I ended my initial editorial 6 years ago by likening the editorship of *CD* to owning a Patek Philippe watch, about which it is said, "You never actually own a Patek Philippe. You merely take care of it for the next generation." I began my term by saying that I was honored to have been entrusted with this scholarly jewel. As I finish my term, I am relieved to report that the watch is still ticking, that I have very much enjoyed wearing it, and that I am looking forward to the feel of the breeze on my bare wrist.

References

Cairns, R. B., & Ornstein, P. A. (1979). Developmental psychology. In E. Hearst (Ed.), *The first century of experimental psychology* (pp. 459–512). Hillsdale, NJ: Erlbaum.

Harris, B. (1979). Whatever happened to Little Albert? *American Psychologist, 34*, 151–160.

Kessen, W. (1965). *The child*. New York: Wiley.

Liben, L. S. (1997, Fall). Standing on the shoulders of giants—Or collapsing on the backs of straw men? *The Developmental Psychologist: The Newsletter of Division 7 of the American Psychological Association*, 2–14.

McGraw, M. B. (1935). *Growth: A study of Johnny and Jimmy*. New York: Appleton-Century.

Merton, R. K. (1985). *On the shoulders of giants*. San Diego, CA: Harcourt Brace Javanovich. (Original work published 1965, New York: Free Press)

Overton, W. F. (2006). Developmental psychology: Philosophy, concepts, methodology. In W. Damon & R. M. Lerner (Series Eds.) & R. M. Lerner (Vol. Ed.), *Handbook of child psychology: Vol. 1. Theoretical models of human development* (6th ed., pp. 18–88). New York: Wiley.

Zeaman, D. (1959). Skinner's theory of teaching machines. In E. Galanter (Ed.), *Automatic teaching: The state of the art* (pp. 167–176). New York: Wiley.

Child Development, November/December 2008, Volume 79, Number 6, Pages 1600–1605

Continuities and Discontinuities in Children and Scholarship

Lynn S. Liben
The Pennsylvania State University

This article introduces a collection of essays on continuity and discontinuity in cognitive development. In his lead essay, J. Kagan (2008) argues that limitations in past research (e.g., on number concepts, physical solidarity, and object permanence) render conclusions about continuity premature. Commentaries respectively (1) argue that longitudinal contexts are essential for interpreting developmental data, (2) illustrate the value of converging measures, (3) identify qualitative change via dynamical systems theory, (4) redirect the focus from states to process, and (5) review epistemological premises of alternative research traditions. Following an overview of the essays, this introductory article discusses how the search for developmental structures, continuity, and process differs between mechanistic-contextualist and organismic-contextualist metatheoretical frameworks, and closes by highlighting continuities in Kagan's scholarship over the past half century.

In his essay, "In Defense of Qualitative Changes in Development," Kagan (2008) addresses a contemporary controversy: Are the cognitive functions seen in infancy preserved into and beyond childhood or are they so transformed in ontogenesis that they should be viewed as qualitatively different functions? Kagan mounts an argument for the need for caution in the resolution of this debate. The five commentaries address Kagan's points and offer additional arguments and empirical data bearing on the continuity–discontinuity issue. Taken together, the articles demonstrate the diverse ways that developmental scientists conceptualize, empirically study, and draw conclusions about continuous versus qualitative change.

Delineating the Questions

Kagan (2008) observes that scientists from a number of disciplines take as their primary goal "to explain how a phenomenon of interest developed its observed form" (p. 1606). He illustrates his point by drawing from evolutionary biology, embryology, and, his focus here, developmental psychology. After providing some historical perspective on prevailing developmental beliefs of earlier eras, Kagan critiques the currently popular view that individual development is best characterized as quantitative rather than qualitative change. Typical of such accounts are nativist theories and research programs that imbue infants with an assortment of underlying and fundamentally adult-like competencies that need merely to be triggered and sharpened by contact with the world. Kagan notes that the theoretical and empirical claims for preservation of function from infancy to maturity have been offered in two major arenas of contemporary developmental psychology, one the preservation or continuity of early emotional attachment, the other the preservation or continuity of early cognitive functions. It is the second of these he addresses in the core of his article.

Kagan identifies two major problems with contemporary research programs that promulgate cognitive continuity. The first concerns language. Kagan argues that too often identical labels are used for behaviors that might, on the surface, appear identical, but that, at root, might be distinct. One illustration he gives is the use of the term *imitation* to label an infant's tongue protrusion in the presence of an adult's tongue protrusion. In the neonate, protrusions might reflect infants' tendencies to use their tongues to explore nearby objects; in the older infant, they might indicate infants' representation-based copying of another person's behavior. Applying the identical term—*imitation*—to both earlier and later instantiations of an observable behavior, "tempts less sophisticated readers to assume that the two phenomena are exemplars of the same process" (p. 1607). Kagan thus urges the use of different terminology for cognitive functions that are what he calls schematic (containing only perceptually available information) from those he calls semantic (containing additional, based

---

I am grateful to Michael Chandler and Willis Overton for their thoughtful comments on an earlier draft of this article.

Correspondence concerning this article should be addressed to Lynn Liben, Department of Psychology, 450 Moore Building, The Pennsylvania State University, University Park, PA 16802. Electronic mail may be sent to liben@psu.edu.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7906-0002

information as well). Of course, just as someone wishing to leave room for qualitative change may argue that identical labels for earlier versus later behaviors are misleading, so, too, someone arguing for continuity may argue that *different* labels are misleading. Despite the possible legitimacy of both positions, the former arguably has more traction because—as in any coding system—one can collapse categories later, but one cannot recover distinctions that were initially ignored. Even then, though, one would still be left with the dilemma of deciding—in the absence of agreement about what underlies any given observable behavior or how to differentiate one observable behavior from another—when to apply which label.

Kagan's second major critique revolves around the empirical database that he claims is far too weak to support conclusions about infants' cognitive functions that are routinely drawn by contemporary investigators. Kagan, like others before him (e.g., Haith, 1998) faults the field for its heavy reliance on looking time methods. His critique is in part based on the varied bases for looking time per se and in part based on the more general point that converging methods and data are needed in research. Kagan draws from research in three domains—the concept of number, physical solidity, and object permanence—to illustrate his conceptual and methodological points related to continuity and transformation. He ends his article by urging the use of a greater range of measures and greater attention to the role of the maturing brain.

Kagan's Gauntlet

Kagan entitled his article a defense, but more accurately, it is a challenge. Indeed, in the opening sentence of their commentary, Campos et al. (2008) remark that Kagan "has dropped a gauntlet, challenging the major intellectual and empirical trends in infancy research of the past quarter century" (p. 1625). The five commentaries respond to the challenge set before them, albeit in very different ways.

Expanding the Challenge

Campos et al. (2008) violently agree with Kagan. They endorse the need for converging evidence and replications and offer additional examples of overinterpretations of empirical data. The bulk of their commentary is devoted to expanding on Kagan's argument, often linking their points to metatheoretical issues (many of which are addressed explicitly in the final commentary by Müller & Giesbrecht, 2008). Their overriding theme concerns the essential need to embed all developmental research within a longitudinal context. Without it, interpretation of a given behavior at a given age may be misleading. They provide several specific examples of behaviors (e.g., imitation and reaching) that appear to be morphologically identical across age but that are revealed to be qualitatively different once one examines the broader context in which the given action is embedded. They also develop the argument that particular behaviors—even those that serve a function at a particular period of the life span—may or may not have consequences for or connections to later competencies. Behaviors may regress as well as progress. Further, they argue that development need not always consist of new individual functions; it can sometimes represent novel forms that emerge by nonadditive reorganizations of earlier functions. However, as becomes clear both within the current set of commentaries as well as from far more extensive discussions of metatheoretical issues in developmental science (e.g., Overton, 2006, 2007; Witherington, 2007), the centrality and scientific legitimacy of the search for novel, emergent outcomes differs across metatheoretical frameworks.

Accepting the Challenge

Quinn (2008) accepts Kagan's challenge as an empirical challenge. Quinn agrees with Kagan that the findings from many empirical studies with infants have indeed been flawed and overinterpreted. He argues, however, that some investigations do use the kinds of empirical approaches to study infant cognition that Kagan recommends. To illustrate, he selects the example of infant categorization and draws from his program of research to illustrate how research using multiple approaches (e.g., looking times, event-related potentials, computational modeling) converge on the conclusion that category structures of infants (e.g., for cats and dogs) are indeed similar to those of adults. Thus, Quinn makes the case for why, to use Kagan's terminology, schematic categorization may indeed be linked, continuously, to semantic categorization. Interestingly, though, the "structural similarity" of Quinn's illustrative focus is on whether infants' and adults' assignments of animals into cat versus dog categories match. It is not, for example, on whether infants and adults share an understanding of the logical relations between and among categories (e.g., cats, dogs, animals) such as would be taken as a hallmark of classification understanding by Piaget (1952).

Redefining the Challenge

Spencer and Perone (2008) join Kagan in his defense of qualitative change, writing from their explicit commitment to dynamical systems theory (DST). Before assuming highly congruent perspectives, however, it would be wise to remember Kagan's first point about problems stemming from the use of identical words for what may be fundamentally different constructs or behaviors. To Spencer and Perone's credit, they explicitly note the confusion that results from discrepant definitions, and they thus begin their commentary by providing their theory-embedded definition of the term "qualitative." Specifically, and as explained in detail in the commentary, the term qualitative change is used within DST to refer to changes from one stable attractor state to another. One feature of their approach is that the behaviors being modeled are unfolding on the second-to-second timescale. Another is that qualitative transformations (as defined within DST) can emerge from continuous changes. Spencer and Perone argue that the study of qualitative shifts over broad periods of time (as in the years between infancy and childhood) show little about the way behavior in real time is organized. Their approach is thus based on examining tiny units that are intentionally removed from the kinds of analyses and constructs promulgated by Campos et al. (2008; see also Witherington, 2007).

Redirecting the Challenge

Adolph and Robinson (2008) offer a perspective that redirects the focus of discussion. They argue that much of the conversation about development has emphasized early versus late states per se, and that scientists interested in progressions would be better served by turning their attention away from characterizing beginning, end, and intermediate states (stages) and toward the *process* of developmental change. They, like Spencer and Perone, ground their work in a systems framework, although the two groups provide different kinds of illustrations. Spencer and Perone illustrate their work by running tens of thousands of computer simulations of neuronal behavior. Their figure 1 shows "real-time changes in the behavior of our neuron (or neural field) over a time scale that might encompass a day or two of experience" (Spencer & Perone, 2008, p. 1644). Adolph and Robinson illustrate their work by addressing the development of walking, a progression that follows a far more protracted path. Although the change from crawling to walking is sometimes used to illustrate qualitative change, Adolph and Robinson instead use it to argue for core continuities: Alternating leg movements are seen during the prenatal period, soon after birth in kicking, later in infancy in crawling, and still later in toddlerhood and beyond in walking. The timescale of this substantive focus is thus far more similar to the kinds of phenomena discussed by Kagan in his lead article. It would be intriguing to try to analyze the structural similarities and differences of a particular motor problem (like alternating leg movements) versus a particular cognitive problem (like a concept of number) in considering how the kinds of methods used by Adolph and Robinson might (or might not) be transferable to the illustrative domains discussed by Kagan.

Contextualizing the Challenge

Müller and Giesbrecht (2008) join Campos et al. (2008) in arguing for the importance (and legitimacy) of examining any particular behavior at any particular moment *within* (rather than isolated from) an extended developmental sequence. Furthermore, they note the value of interpreting earlier steps in view of final steps, citing Piaget's method of recurrence as a model. This approach contrasts strikingly with the second-to-second methodology of DST discussed by Spencer and Perone (2008), which is intentionally focused on the "local details of behavior." The bulk of Müller and Giesbrecht's commentary is directed to the broader epistemological differences between contemporary neonativist approaches that emphasize continuity across age (e.g., Baillargeon, Spelke, Wynn) and classic theories that identify transformations across age (e.g., Piaget, Werner, Vygotsky). They situate their discussion of programs of empirical work within the philosophical traditions of rationalism and empiricism, explain ways in which different epistemological premises lead different theories to pose fundamentally different questions and, relatedly, to use fundamentally different methods. Müller and Giesbrecht illustrate these points by contrasting nativist versus constructivist accounts of both representation and reasoning, drawing specific contrasts between the research of Baillargeon and Piaget.

At the End of the Joust

Jousts are categorized as *à plaisir* (for pleasure) or *à l'outrance* (to the death). The joust ensuing from Kagan's tossed gauntlet surely has not resulted in the death of either the continuity or the discontinuity positions. What, then, can be concluded from this collection of articles?

First—as argued in virtually any article addressing virtually any domain—the field needs more and better empirical work. Particularly important are programs of research that bring converging methods to bear on a given conclusion. This point is made emphatically and clearly by Kagan in his lead article, and it is endorsed in all five commentaries. Campos et al. (2008) provide some additional principles for how empirical work can be better; Quinn (2008) provides specific illustrations of better empirical work that can serve as a model for converging designs. Spencer and Perone (2008) as well as Adolph and Robinson (2008) discuss alternative methodologies and analytic tools for new kinds of empirical research.

Second—and again as recommended in virtually all developmental articles—the field needs more and better research on developmental process and not just on products. This point is made most explicitly in the two commentaries by Adolph and Robinson (2008) and by Spencer and Perone (2008). Observing and modeling behaviors as they unfold over time can reveal trajectories and patterns that would otherwise be invisible in the behavior itself. At the same time, though, the collection of commentaries makes it clear that recording and modeling observable incremental behaviors does not in itself resolve the problem of identifying process. That is, one can document and model behavioral differences at increasingly more microscopic levels of analysis, but this alone need not reveal causal mechanisms, at least not the kinds of causal mechanisms that may be sought. An illustration of this issue is found in Müller and Giesbrecht's (2008) discussion of the contrast between the way that Baillargeon (2008) and Piaget (1963) approach infant reasoning. They argue that Baillargeon endows the infant with nativist principles or computational devices that process encountered information. Given these devices, the infant is in essence merely a spectator. In contrast, Piaget endows the infant with the functional activity of assimilation. This assimilation (and its conjoined twin, accommodation) entails the infant's active engagement with the world, thus rendering the infant an intentional agent. But are both kinds of processes scientifically legitimate or even real? The answer to this question is, "It depends." That is, an explanatory process that is legitimate within one theory (e.g., assimilation and accommodation of Piaget) is dismissed as ephemeral by another (as when Klahr labeled these the "Batman and Robin" of developmental psychology; see Adolph & Robinson, 2008). These conflicting answers lead directly to the last point (which in essence encompasses the prior two).

The third message that may be derived from these articles is that the field needs to attend more actively to the metatheoretical context in which research is designed, conclusions are drawn, and disagreements are discussed. A failure to contextualize theories and empirical data in metatheory results in scholars misconstruing one another's meaning, or, alternatively, ignoring one another's meaning altogether. This point is developed most explicitly by Müller and Giesbrecht (2008), but it pervades the collection of articles. Constructs (e.g., concept of number, cat and dog categories) are formulated in radically different ways depending on the metatheoretical frameworks from which they emerge; likewise, the search for origins and processes differs.

As discussed in detail by Overton (2006, 2007) and Witherington (2007), theories rooted in the principles of a mechanistic – contextualist framework necessarily assert the existence of a base level of pure elements (e.g., behaviors, neurons, genes) that combine additively to produce wholes. Hence, the strategy of mechanistic – contextualist approaches is to study ever finer, directly observable units, with the goal of reaching a foundation that reveals chains of cause and effect elements. This is the mechanistic – contextualist interpretation of "process." In contrast, theories rooted in an organismic – contextualist framework eschew the assumptions of additivity of elements and the ultimate goal of identifying a bedrock foundation. Instead, their holistic perspective asserts that truly novel forms or dynamic systems emerge as the system acts upon the world. What emerges is thus different from the sum of the parts; integration yields novel systemic properties that are not characteristic of any of the parts. There is still continuity insofar as later forms subsume earlier forms, but there is also discontinuity or qualitative change.

To take a perhaps worn but still useful analogy from chemistry, the viscosity, color, and drinkability of water are emergent characteristics of water that could not be predicted from or observed within the properties of its constituent hydrogen and oxygen. Explicating the moment-by-moment processes by which electrons from hydrogen and oxygen merge in the outer rings of the molecules would still fail to explain or predict the qualities of water. Kagan's example of the movement from a perceptually based representational system to a symbolically based representational system is a more immediate developmental example. Within the mechanistic – contextualist perspective, empirical work entails the search for the bedrock appearance of the symbol at birth (or perhaps later, but still merely triggered by some biological or environmental stimulus). In the contrasting organismic–contextualist perspective, empirical work entails identifying an emergent property of a system arising from actions and integrations of earlier sensorimotor action systems. Here, symbolic representation would not be expected to be apparent in earlier forms, any more than qualities of water would be expected to be apparent in hydrogen and oxygen. Neither approach will be validated by closer observations or by converging data. Their adequacy will be assessed according to the criteria that are generally applied in the selection of any adequate scientific explanation such as the explanation's depth, coherence, and logical consistency.

In the end, then, it should be clear that the current collection of articles does—as claimed in the initial paragraph of these introductory comments—"demonstrate the diverse ways that developmental scientists conceptualize, empirically study, and draw conclusions about continuous versus qualitative change." It should be equally clear, though, that these thoughtful contributions do not converge on a single position. As long as scholars approach development from different epistemological frameworks and metatheories, there will be plurality, not consensus.

Continuity in Scholarship on Continuity

Irrespective of one's position on continuity in the study of individual development, in the everyday meaning of the word, it is incontrovertible that there is continuity in Kagan's scholarship. Over half a century ago, Kagan and Moss (1962) opened their seminal book on the Fels longitudinal study by saying: "The search for stability in physical and biological systems is a lure that attracts much scientific inquiry. The stability of that system we popularly call human behavior has long whetted the curiosity of most men, be they scientists *qua* scientists or scientists *qua* parents" (p. 1). Save for the now anachronistic use of the word *men* rather than *people*, this opening sentence could have been pulled from Kagan's current article.

And, roughly two decades after the appearance of the Fels book, but still roughly three decades ago, Kagan coedited with Brim a collection on *Constancy and Change in Human Development* (Brim & Kagan, 1980a). What made that volume so novel at the time was first, its consideration of stability and change across the entire life span, and, second, its incorporation of non-Western concepts of stability and change. The chapter by Kagan (1980) within it provided a rich discussion of how the concept of continuity has been conceptualized from a range of disciplines, eras, and philosophical positions.

Despite the long history of work on continuity by Kagan and others, key questions remain unresolved. As discussed in the commentary by Müller and Giesbrecht (2008), and as amplified in both earlier and contemporary analyses of these issues (e.g., Kagan, 1980; Overton, 2006), these questions cannot be resolved simply by better and better empirical work. In this context, it is useful to recall Brim and Kagan's (1980b) early observation: "Constancy and change are inventions of the mind, of course, a matter of subjective magnitude" (p. 2).

Given that readers will approach the articles that follow with different theories and metatheories, it is unlikely that all will derive the same conclusions about continuity and discontinuity from reading them. All will, however, come away from this collection thinking more deeply and productively about the rich conceptual and empirical work on continuity already in our books and journals, and better equipped to think more deeply and productively about the related work that will undoubtedly continue to fill our literature in the decades to come.

References

Adolph, K. E., & Robinson, S. R. (2008). In defense of change processes. *Child Development*, *79*, 1648–1653.

Baillargeon, R. (2008). Innate ideas revisited: For a principle of persistence in infants' physical reasoning. *Perspectives on Psychological Science*, *3*, 2–12.

Brim, O. G., Jr., & Kagan, J. (1980a). *Constancy and change in human development*. Cambridge, MA: Harvard University Press.

Brim, O. G., Jr., & Kagan, J. (1980b). Constancy and change: A view of the issues. In O. G. Brim Jr. & J. Kagan (Eds.), *Constancy and change in human development* (pp. 1–25). Cambridge, MA: Harvard University Press.

Campos, J. J., Witherington, D., Anderson, D. I., Frankel, C. I., Uchiyama, I., & Barbu-Roth, M. (2008). Rediscovering development in infancy. *Child Development*, *79*, 1625–1632.

Haith, M. M. (1998). Who put the cog in infant cognition? Is rich interpretation too costly? *Infant Behavior & Development*, *21*, 167–179.

Kagan, J. (1980). Perspectives on continuity. In O. G. Brim Jr. & J. Kagan (Eds.), *Constancy and change in human development* (pp. 26–74). Cambridge, MA: Harvard University Press.

Kagan, J. (2008). In defense of qualitative changes in development. *Child Development*, *79*, 1606–1624.

Kagan, J., & Moss, H. A. (1962). *Birth to maturity: A study in psychological development*. New York: Wiley.

Müller, U., & Giesbrecht, G. (2008). Methodological and epistemological issues in the interpretation of infant cognitive development. *Child Development*, *79*, 1654–1658.

Overton, W. F. (2006). Developmental psychology: Philosophy, concepts, methodology. In W. Damon & R. M. Lerner (Series Eds.) & R. M. Lerner (Vol. Ed.), *Handbook of child psychology: Vol. 1. Theoretical models of human development* (6th ed., pp. 18–88). New York: Wiley.

Overton, W. F. (2007). A coherent metatheory for dynamic systems: Relational organicism-contextualism. *Human Development*, *50*, 154–159.

Piaget, J. (1952). *The child's conception of number.* London: Routledge and Kagan.

Piaget, J. (1963). *The origins of intelligence in children.* New York: Norton.

Quinn, P. C. (2008). In defense of core competencies, quantitative change, and continuity. *Child Development*, *79*, 1633–1638.

Spencer, J. P., & Perone, S. (2008). Defending qualitative change: The view from dynamical systems theory. *Child Development*, *79*, 1639–1647.

Witherington, D. C. (2007). The dynamic systems approach as metatheory for developmental psychology. *Human Development*, *50*, 127–153.

Child Development, November/December 2008, Volume 79, Number 6, Pages 1606–1624

In Defense of Qualitative Changes in Development

Jerome Kagan
Harvard University

The balance between the preservation of early cognitive functions and serious transformations on these functions shifts across time. Piaget's writings, which favored transformations, are being replaced by writings that emphasize continuities between select cognitive functions of infants and older children. The claim that young infants possess elements present in the older child's concepts of number, physical impossibility, and object permanence is vulnerable to criticism because the inferences are based primarily on the single measure of change in looking time. It is suggested that investigators use unique constructs to describe phenomena observed in young infants that appear, on the surface, to resemble the psychological competences observed during later developmental stages.

The primary goal of scientists working in varied disciplines is to explain how a phenomenon of interest developed its observed form. A central theoretical issue in this search centers on the degree to which the current event contains some features that were present earlier and contribute to the contemporary observation. Darwin's great insight was to recognize that many of the anatomical features and functions of mammals were either absent or assumed a very different pattern in the organisms that lived millions of years earlier from which they presumably evolved. For example, the convoluted cortex of humans is not present in fish, and although the three bones of the human middle ear are derivatives of a set of bones in the fish jaw, they have a qualitatively different structure and function in the two species. Thus, it is correct to say that the bones of the fish jaw preceded the bones of the middle ear but incorrect to declare that the functions and features of the former structure contribute to human hearing.

The principle that some elements are preserved and others are transformed or lost applies to developmental stages within a species. On the one hand, the glial cells of the young embryo that guide newly created neurons to various cortical layers disappear once the mature neurons arrive at their final destination (Rakic, 1995), and the motility patterns of newly hatched birds pecking at the ground have no parallel in the motor profiles of the developing embryo (Hamburger, 1975). On the other hand, many structures and functions of the embryo's heart and mouth, once they appear, are preserved indefinitely.

The balance between transformation and preservation, which has always been a seminal node of controversy among developmental psychologists, has shifted over time as a function of new technologies, discoveries in allied disciplines, the need to replace a flawed theoretical view, and always the intellectual climate. Many social scientists had attributed the differences in academic achievement between the children of early 20th-century European immigrants to America and those of more affluent, indigenous families to the preservation of early habits and skills acquired within the family. A 1914 government pamphlet for mothers contained a warning that, although it rested on little evidence, reified the conviction that first impressions can have profound influences on the future: "The first nervous impulses which pass through the baby's eyes, ears, finger, or mouth to the tender brain make a pathway for itself; the next time another impulse travels over the same path, it deepens the impression of the first" (West, 1914, p. 59).

This belief in the preservation of the products of early experiences, which was central to behaviorism and psychoanalytic theory, thrived for over 50 years. However, the emergence of a new point of view was inevitable when the accumulated evidence failed to support strong versions of this hypothesis. Ethologists, especially Lorenz (1958) and Tinbergen (1951), who anticipated the ascent of biological mechanisms, recruited interest to the biologically prepared competences of young organisms that might be components

---

I thank Richard Aslin, Susan Carey, Leslie Cohen, Marshall Haith, David Moore, Charles A. Nelson, Nora Newcombe, J. Steven Reznick, and Jay Schulkin for comments on an early draft. The incentive for this article was a dinner conversation with Michael Lewis at the 2007 meeting of Society for Research in Child Development in Boston.

Correspondence concerning this article should be addressed to Jerome Kagan, Department of Psychology, Harvard University, Cambridge, MA 02138. Electronic mail may be sent to jk@wjh.harvard.edu.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7906-0003

of later functioning. Most new scientific ideas require between 40 and 50 years before they become popular bases for inquiry. By the late 1990s, technical reports describing biologically prepared conceptual functions in young infants that shared features with those of older children began to fill journal pages (Baillargeon, 1994; Dehaene, 1997; Saxe & Carey, 2006; Spelke, 1994; Spelke & Kinzler, 2007; Wynn, 1992).

When understanding of a prolonged developmental cascade is lean, and it is difficult to imagine the intervening stages, there is a strong temptation to assume the smallest possible number of transformations between early and later phenomena in order to be loyal to the principle of parsimony. That is one reason why the 18th-century preformationists assumed that a miniature infant was present in the newly fertilized egg and why Havelock Ellis perceived a link between the look of satisfaction on the face of a nursing infant and the facial expression of lovers lost in the ecstasy of their sexual caresses (Ellis, 1900). Our fragile understanding of the degree to which the behavioral, cognitive, and emotional features of the young infant participate in more mature forms tempts theorists to assume that the early representations and functions are both necessary for, and contain some features of, those that characterize the psychology of school-aged children.

However, the habit of applying the same constructs to explain early and later performances, based on presumed similarities between the bases for early and later behaviors, does not guarantee that the initial and later forms were mediated by the same mechanism. The presence of a reliable smile to an approaching adult, which usually emerges between 3 and 5 months, reflects the fact that the infant is now able to assimilate the person's features to their schemata for the adults with whom they interact regularly. It is unlikely that this mechanism is the reason why an adolescent smiles upon greeting a close friend. Two-week-olds and 2-year-olds are likely to protrude their tongue following the same act by an adult (Meltzoff & Moore, 1977), but young infants also protrude their tongue to a slim pencil approaching their mouth because infants use their tongue to explore objects close to their faces (Jacobson, 1979; Jones, 1996; Nakagawa, Sukigara, & Benga, 2003). Hence, if the older child's behavior is called imitation, a new term should be invented to describe the apparently similar behavior of the young infant. Earlier generations of psychologists assumed that scores on the test items of the Bayley developmental scale for infants and those on the Wechsler Intelligence Scale for school-aged children represented the same psychological construct, called *intelligence*. However, the magnitudes of the correlations between the Bayley scores at 3 or 8 months and the IQ scores of older children are minimal (McCall, Eichorn, & Hogarty, 1977); hence, the infant competences should be given a different name.

The uncritical use of the same semantic term to name early and later behavioral profiles that are phenotypically distinct and that might rest on different foundations may tempt less sophisticated readers to assume that the two phenomena are exemplars of the same process. These readers should appreciate that current evidence has not yet proven with certainty that any representation, emotion, or cognitive competence present in healthy 6-month-olds is contained in the psychological functions of older children. The mind possesses a small number of stubborn biases that color the initial perception of events and the conception of the hidden structures and functions that lie beneath what is observed. One of the most profound biases, which children honor, is that "things" possess an essence that is unchanging despite surface malleability. Although young infants possess some sensory, perceptual, and motor abilities that are preserved indefinitely (e.g., the detection of sweet tastes and sounds and an automatic orienting of the eyes and head to a moving stimulus), some scholars have posited cognitive competences that go beyond these talents to include functions that are more seriously symbolic or conceptual in nature (Baillargeon, 1994; Spelke, 1994; Wynn, 1998). This claim deserves careful analysis because, if valid, it has the potential to alter our understanding of the infant and development in a radical way.

This issue assumes two major forms in contemporary research: the preservation of early emotional attachments to caretakers and the preservation of early cognitive functions. Discussion of the former has been substantial (Kagan, 1981). Hence, this article discusses only the latter issue and continues a critique initiated by others (Bogartz, Shinskey, & Schilling, 2000; Bogartz, Shinskey, & Speaker, 1997; Cashon & Cohen, 2000; Cohen & Cashon, 2006; Haith, 1998; Haith & Benson, 1998; Mandler, 2004).

The nub of the controversy centers on how much (or how little) transformation of the infant's psychological structures occurs as the child grows. If the events that transpire after 6–8 months alter early structures only quantitatively, the consequences of lawful brain growth and later experiences would seem less interesting than they would be if these changes produced qualitative transformations on the original functions. There is an important difference in meaning between (a) "The newborn's fingers closed on the object placed in its palm" and (b) "The newborn possesses a disposition for possessiveness." Many 19th-century observers cited the newborn's reflex reaction as the basis for the validity of the second statement (Fiske, 1883).

The histories of scientific controversies reveal that when two groups of sophisticated investigators disagree on the interpretation of a corpus of evidence, later inquiry usually indicates that the most fruitful explanation differs from the one promoted by either group. Early 20th-century physicists argued over whether light was a wave (the classical view) or a particle (the new quantum position). Bohr suggested that physicists should not ask what light "is" but rather how light behaves under specific experimental conditions because the experimental arrangement determines whether a wave or a particle interpretation is the most defensible. The consensus among neuroscientists working in the 1920s was that all synaptic transmission in the brain was electrical in nature. This view was challenged by a small group claiming transmission was chemical. We now appreciate that although most synapses are chemical, both mechanisms occur (Valenstein, 2005). This issue is important enough to return to it again with a different organization of facts and from a slightly different point of view.

Schemata and Semantic Forms

The psychologists who remain skeptical of the nativist claims for preservation of select early cognitive functions believe that the young infant's knowledge is contained in schemata and sensorimotor structures, whereas the older child's knowledge combines these two earlier forms with semantic networks and the ability to infer events that were not experienced. This suggestion implies a discontinuity in cognitive development because the features and organizations of the early functions differ seriously from those operative later (Cohen & Cashon, 2006; Haith & Benson, 1998; Kagan, 2002; Mandler, 2004). A schema is defined as a representation of the patterned perceptual features of an event and need not possess a semantic component or an inference as to its origin. The young infant's representation of a caretaker's face is an obvious example. The primary features of schemata for visual events are spatial extent, shape, color, contrast, motion, symmetry, angularity, circularity, and density of contour.

Eleven-year-old Mayan Indian children living in a small isolated, impoverished village in northwest Guatemala performed as well as 11-year-old Boston children when asked to recognize, following a 2-day delay, a large number of color photographs of unfamiliar objects for which they had no names, such as telescope, toaster, and golf club. Both groups of children correctly recognized the pictures close to 90% of the time (Kagan, Klein, Haith, & Morrison, 1973). It is likely that the excellent memory performance of the Mayan children was mediated by the schemata for the pictures they acquired during familiarization.

Infants, like older children, create schematic concepts or prototypes that are "averages" of repeated exposures to the same class of events. Eight-month-olds have already acquired schematic concepts for the acoustic envelopes of the vowels and consonant–vowel combinations of their language. However, they lose the ability to distinguish between phonemes that resemble the speech sounds they hear but do not occur frequently (Doupe & Kuhl, 1999). The same phenomenon occurs for schemata representing the mouth movements of speakers (Weikum et al., 2007). Infants can also create schematic concepts for temporal patterns of unfamiliar vocal sounds. Seven-month-olds first heard a 2-min speech sample containing three representations of each of 16, three-syllable utterances of the form a-b-a, but the syllables in each of the utterances were different. The only feature shared by all the utterances was that the first and third syllables were identical. After the infants were familiarized with this pattern, they heard either an utterance of the same form—a-b-a—or a new set of three syllables with a new form; for example, a-b-b. The infants were more attentive to the unfamiliar a-b-b pattern, suggesting that they had created schematic concepts for the temporal pattern of these meaningless vocal sounds (Marcus, Vijayan, Rao, & Vishton, 1999).

An important distinction between schemata and semantic structures is that only the latter imply both the essential features a referent possesses and some features it does not possess. This logic does not apply to schemata. The sentence, "The leaf is on the ground," implies that there was a time in the past when it was not on the ground; the schema of a leaf on a lawn does not contain that extra knowledge. A second distinction is that most of the time schemata can be transformed without any accompanying cognitive tension. Dreams and reveries provide classic examples. The image of a smiling face can turn into a pumpkin without any of the cognitive dissonance that would occur if a person read, "A smiling face is a pumpkin." Einstein's image of riding a light wave is said to have been one of the origins of his theory of special relativity. However, the printed sentence, "Humans can move at the speed of light" would evoke in most readers an immediate sense of impossibility.

An important difference between schemata and semantic forms involves their relations to other structures. The relations of similarity, part–whole, and contingency apply to both schemata and semantic forms, even though the specific attributes that define these relations are different for the two structures. Some relations, however, apply only to semantic structures. Only semantic forms possess the quasi-logical relations that take the form of hierarchically nested categories for objects and events (called hyponyms) and opposites for qualities (antonyms). These relations assume special prominence after 5 or 6 years of age. Most adults treat the semantic form "dog" as a logically necessary member of the semantic category "animal." Schemata do not nest into sets of hierarchical categories.

Semantic structures vary in their schematic contribution; hence, psychologists compare concrete with abstract nouns. Words like *meaning* and *truth* have far less schematic content than *girl* or *table* (Warrington & McCarthy, 1987). The different properties of semantic and schematic representations are due, in part, to the differential participation of the right and left hemispheres. The right hemisphere appears to be biologically prepared to process words with a rich schematic contribution more efficiently than it processes abstract words (Beeman & Chiarello, 1998; Deloche, Seron, Scius, & Segui, 1987; Ivry & Lebby, 1998). The meaning of the semantic term *big* depends on the specific schemata that are activated. If the child retrieved a schema for trains, big meant long. If, however, the child activated the schema for pumpkins, big meant a large area, and if the activated schema was a building, big meant tall (Sera & Smith, 1990). The semantic meaning of big varied with the schemata the child activated.

Schemata are used primarily to recognize events, places, and objects experienced in the past and often include representations of the background contexts in which events appear (Moore & Meltzoff, 2004). By contrast, semantic representations, which are often free of contextual information, are used primarily for classification, inference, comprehension of communications, and understanding causal sequences. The critical point is that schemata and semantic representations for an event award have different meanings to the predicate "know." Indeed, infants who normally treat a picture as discrepant from a familiarization set do not do so when they hear the same verbal label (a nonsense word) accompany the presentation of all stimuli during the familiarization trials (Plunkett, Hu, & Cohen, 2008). This observation implies that combining semantic information with schemata can mute the physical differences between the perceptual representations. The act of naming apples, oranges, and kiwi fruit dilutes, at least for the moment, the mind's awareness of the perceptual distinctions among these objects.

The differences between schemata and semantic forms are sufficiently salient to warrant asking whether the infant's ability to perceive the difference between arrays of two and four circles reflects the same competence that underlies the ability of 10-year-olds to learn the names for cardinal numbers and to appreciate that they form an ordinal scale of magnitude. A schema for an event, or schematic prototype for a category of event, can exist without any support from semantic elements; hence, the former are to be distinguished from semantic concepts. I would be surprised, and stare for a few seconds, if the number of books on the shelf several feet in front of my desk (and in my direct line of vision) were altered because the new display would violate my schema for that scene. However, I have no idea of the number of books on the shelf. Children watching a video of two chromosomes separating during mitosis will create a perceptual representation for the sequence without any conceptual understanding of what is happening. Young infants and many animal species can discriminate between pictures of an elm tree and an elephant and between parallel and nonparallel lines. But this evidence does not mean that they possess the semantic concepts plant, animal, or parallel.

It is also the case that individuals can have some semantic understanding of a phenomenon without a corresponding set of schematic representations. I know that if I were bitten by an Anopheles mosquito infected with malarial parasites I would come down with a fever, even though I have no schemata for the parasite or the toxins it produces in my body. The extensive literature showing that infants under 8 months display increased attentiveness to events that violate a perceptual expectation does not require positing any semantic representation implying an understanding of the event's origins or consequences. This suggestion does not mean that young infants do not possess semantically based conceptual representations, only that this hypothesis is not yet proven beyond reasonable doubt (Aslin, 2000).

The Source of Evidence

The meaning and validity of every scientific conclusion based on empirical data always rest on its source of evidence. This principle is one of the most important dividends of the 20th-century advances in physics that culminated in quantum mechanics. The statement that turtles and crocodiles are phylogenetically distant is true if similarity in their skeletons is the source of information but false if similarity in DNA strings is the evidence.

This principle is relevant to the concerns of this article because the evidence for most claims about infants' conceptual capacities is total time attending to particular events rather than other indexes of attention, such as duration of the first fixation, the longest fixation, the number of times the infant looks away from the stimulus, and the number of shifts of attention between a pair of events. Aslin (2007) notes that these variables might reflect different cognitive processes, a position articulated by Lewis (1967) over 40 years ago. This possibility is supported by the fact that infants younger than 1 year who displayed frequent attention shifts between a pair of stimuli, together with shorter individual fixations, showed larger increases in attention to a novel event than others (Frick & Colombo, 1996; Rose, Feldman, & Jankowski, 2003). In addition, the number of shifts between a pair of stimuli distinguished between drawings with different element densities better than total fixation time (Tan & Bryant, 2000) and the mean and the longest fixation time often invite different conclusions (Courage, Reynolds, & Richards, 2006).

In addition, most investigators probing infants' concepts do not quantify brain profiles, autonomic reactions, conditioned operants, facial expressions, crying, vocalizing, or smiling as possible indexes of the state provoked by violation of an expectation. The essential premise in the design of most experiments is that unexpected events recruit longer bouts of attention than those that are expected. These investigators assume implicitly that events that do not match an acquired perceptual representation create a state that neuroscientists believe is accompanied by activation of the parahippocampal region, amygdala, and parietal lobe and, as a consequence, a prolonged orientation (Joseph et al., 2006). However, I will orient and stare out the window of my study for several seconds if I hear an unexpected outside sound and will gaze at the faucet in my kitchen if no water flows when I turn it, without any understanding of the origin of the sound or the reason for the failure of water to flow. A prolonged orientation to an unexpected event is not synonymous with knowing or inferring, anything about its origins or conceptual category (Saffran, 2001). Rather, looking reflects a temporary state of uncertainty. That is why 6- and 9-month-olds look at their mother when an unexpected event occurs (Walden, Kim, McCoy, & Karrass, 2007).

Moreover, failure to show a longer bout of attention to one event compared with another does not mean that the infant did not detect the distinctive features of the two events or did not experience an expectancy violation. Infants familiarized on a circle enclosing two dark, horizontally placed circles in the upper part of the design showed increased looking if the horizontally placed circles were vertical but did not show prolonged looking if the circles were replaced with squares, even though infants can detect the difference between squares and circles (Kagan, Linn, Mount, & Reznick, 1979). Thus, failure to show increased attention and the display of increased attention are ambiguous and do not demand only one interpretation.

A Trio of Influences on Looking

Many investigators ignore three factors that can influence duration of attention to an event: (a) the possibility that the relation between unfamiliarity and duration of looking is curvilinear rather than linear, (b) the influence of brain maturation (i.e., uniform changes in all healthy infants in brain structure and physiology over time), and (c) the significance of the physical features of the stimuli.

Linear or Curvilinear Functions

Although the developmental scientists who rely on total duration of attention never say so explicitly, they assume implicitly that the relation between expectation and looking time is linear. The less expected an event (and, by inference, the more discrepant it is from the infant's schemata), the longer the bout of attention devoted to it. Although familiarity and expectedness are not synonyms, it is reasonable to argue that infants and adults have a higher expectation of seeing a familiar than a less familiar event (Cycowicz & Friedman, 2007).

One problem with the assumption that unexpected events always recruit more prolonged attention than expected ones is that the infants' behaviors do not always conform to that prediction. For example, 4-month-old infants who saw a photo of their mother alongside a photo of an unfamiliar woman looked longer at the more familiar parent. They also looked longer at pictures of faces that had the familiar, ecologically valid arrangement of facial features than at faces with the features rearranged, even though the latter were less familiar and therefore less expected. By contrast, 1-year-olds devoted longer bouts of attention to a rearranged face than to a familiar face because their schemata for faces had become much firmer and, as a result, the rearranged faces were now optimally discrepant (Kagan, 2002). Seven-month-old infants looked longer at the vocal source of short sentences containing words they had been familiarized with a day earlier than at the source of spoken sentences containing totally unfamiliar words (Houston & Jusczyk, 2003), and looked longer at a face whose perceptual form matched a vocalization than at a face paired with a discordant vocal sound (Soken & Pick, 1999). These data suggest that the relation between expectation and duration of attention is often curvilinear, rather than linear, when the event is simple, not part of a complex sequence, and there is a minimal burden on the infant's working memory.

Even newborn infants honor this principle. Samples of newborns were familiarized with the nonsense word *titi* played from one of the two speakers placed on either side of the infant's head. The newborns turned their head to the initial presentation of the word. After habituation, the infants heard either the same word or one of the four variations created by altering the fundamental frequency of the original sound by 7%, 14%, 21%, or 28%. The infants were most likely to orient to a stimulus that altered the fundamental frequency by 14% or 21% (moderate discrepancies) and less likely to orient to the more extreme alteration (Weiss, Zelazo, & Swain, 1988; see also Fenwick & Morrongiello, 1998; Saffran, Loman, & Robertson, 2000). Three-week-olds looked longer at 2 × 2 checkerboards than at 8 × 8 and 24 × 24 designs; 8-week-olds looked longest at 8 × 8 checkerboards; and 14-week-olds stared longest at the 24 × 24 forms (Brennan, Ames, & Moore, 1966).

The curvilinear relation between duration of attention and discrepancy is found in older infants. Seven-month-olds were first familiarized with either a sphere or a cylinder. When infants struck a bright yellow padded lever, they were allowed a 2.5-s view of either a sphere or a cylinder. After the infants had habituated and stopped striking the lever, they saw one of the four stimuli. The infants who had been familiarized on the sphere saw a pear-shaped object, club, cylinder, or irregularly shaped styrofoam object. The infants who had been familiarized on the cylinder saw the club, pear-shaped object, sphere, or styrofoam object. The profiles of attention revealed a curvilinear relation between degree of discrepancy from the acquired schema and duration of attention. Infants who had been familiarized on the sphere showed their longest bout of attention to the club, which was moderately discrepant from the sphere; the infants who had habituated on the cylinder showed their longest bout of attention to the sphere. Few infants showed their longest bout of attention to the styrofoam object, even though it was maximally discrepant and, therefore, minimally expected (Hopkins, Zelazo, Jacobson, & Kagan, 1976; see Saffran & Thiessen, 2003 for affirmation of this idea with multisyllable utterances).

McCall and McGee (1976), who examined many studies on the relation between discrepancy and attention, concluded, "There are at least seventeen separate samples of infants that displayed the inverted U-discrepancy function. Seventeen exemplars of one specific quadratic curve would appear to make the discrepancy hypothesis plausible and worthy of further examination" (p. 13).

Maturation

The term *maturation* refers to a uniform progression of changes in brain organization and function in infants growing in any reasonably natural environment. Some of these changes are detailed in a later section. The point to be made here is that the maturational changes that occur between 2 and 3 months affect the relation between looking time and familiarity because infants 4 months or older are more likely than younger infants to relate an event to an established schema. Two-, 4-, and 6-month-old infants sat across from an unfamiliar woman who either interacted with the infant as a parent would or sat with an immobile face displaying a neutral, happy, or sad expression. Because the less mature 2-month-olds did not relate either event to their acquired schemata they looked equally long at both events. The 4-month-olds, who did relate the events to their schemata, devoted longer attention to the stranger when she was behaving like a parent because this posture was moderately discrepant from their past experience; whereas, the immobile posture was too novel and an attempt at assimilation was inhibited (Rochat, Striano, & Blatt, 2002). Two-month-olds are also too immature to perceive an object partially hidden by an occluder as a unitary entity (Johnson et al., 2003) and behave as if they do not expect an object in midair, with no visible support, to fall to a surface (Sitskoorn & Smitsman, 1995). Children will revise that belief when they learn later about astronauts taking space walks.

The maturational changes that occur between 6 and 10 months are accompanied by the introduction of a new process governing attention. Kagan, Kearsley, and Zelazo (1978) demonstrated, over 25 years ago, that holding the stimuli constant, there was a U-shaped function between age and duration of attention across the interval from 4 to 12 months, with a nadir between 6 and 9 months, followed by increased attention to the same events for the rest of the 1st year. Courage et al. (2006) replicated this result with eight different categories of visual stimuli. Other evidence suggests that the enhanced ability to hold a schema for a past event in a working memory circuit and to relate a present event to that schema and to representations acquired in the past are critical additions to the infant's talents at this time (Diamond, 1990; Kagan & Herschkowitz, 2005). As a result, infants older than 8 or 9 months are likely to create a perceptual prototype for frequently experienced events and to ignore discrepancies in single features that do not alter the holistic, gestalt quality of the schema. By contrast, infants less than 8 months often treat discrepancies in single features more seriously (Doupe & Kuhl, 1999; see also Weikum et al., 2007).

The Role of the Stimulus

Many experiments are insufficiently attentive to the physical features of the stimuli presented to infants because they are not of primary interest to the scientist. Ross and Dannemiller (1999) note that motion, color contrast, and luminance compete for the young infant's attention because the brain always reacts to the physical features of an event independent of its meaning. For example, removing or muting the contrast between the color of the background and the color of the target stimulus can affect looking behavior in a significant way. Arrays containing stimuli with red features usually recruit more attention than those with green or blue features, but investigators often ignore this fact when they select the stimuli for their experiments (see e.g., Shimizu & Johnson, 2004). Adults viewing curved or angular versions of neutral objects showed greater blood flow (implying activation) to the amygdala, superior temporal gyrus, anterior cingulate, and select visual areas to the angular compared with the curved forms (Bar & Neta, 2007). Thus, infants might look longer at a toy block than at a hand after seeing a beanbag move spontaneously (without obvious provocation) because the block contained more linear, angular features than the hand and, therefore, might have activated brain sites that provoke a prolonged bout of attention. (See Saxe, Tzelnic, & Carey, 2007, who did not acknowledge this possibility.)

The Concept of Number

I now consider in more detail the evidence behind three classes of claims that do not differentiate between schematic and semantic representations, fail to acknowledge the importance of the source of evidence, or are indifferent to the physical features of the stimulus arrays. The three claims refer to young infants' possession of the concept of number, physical impossibility, and permanent objects. The concept of number is considered first. Some scientists attribute to young infants a concept of number, as contrasted with a perceptual representation of quantity because they can distinguish between displays of different numbers of dark circles in a habituation – dishabituation design. However, Fantz and Fagan (1975) noted, over 30 years ago, that young infants look longer at a design with 20 squares than at one containing only 4 squares because of differences in the amount of contour derived from a combination of the size of each element and the number of elements.

Mix, Huttenlocher, and Levine (2002), who reviewed the literature on the concept of number, concluded that the infant's ability to discriminate among different numbers of elements, as long as the numbers were less than four or five, was based primarily on perceptual features, especially total area, size, and density of the elements, and total contour length (see Tan & Bryant, 2000, for a similar conclusion). When these physical properties were controlled, 6-month-olds behaved as if they could not discriminate between one and two objects (Xu, Spelke, & Goddard, 2005). These observations imply that the infant's ability to perceive differences in the size and density of elements and amount of contour is qualitatively different from the older child's symbolic concept of number (see Feigenson, Carey, & Spelke, 2002). Indeed, in some contexts, 12-month-olds behave as if they can only create and hold in working memory perceptual representations for up to three objects (see Ross-Sheehy, Oakes, & Luck, 2003). One-year-olds who see an examiner place four crackers in a box do not persevere in searching for additional crackers after retrieving the first one, but will do so if they see the examiner place only two crackers in the box, implying that they regard any quantity of cracker greater than three as belonging to a single, amorphous category (Feigenson & Carey, 2005). Le Corre and Carey (2007) suggest that the representation of the cardinal numbers that children use when they count is not an obvious derivative of the infants' perceptual representations of quantity.

My front lawn has eight trees, a fact I was unaware of until moments ago when I counted them. But when I gaze out the window in summer, I do not see a set of eight objects, but an array of tall green-gray objects in the context of a sloping, green lawn. Major changes occur between the infant's perceptual ability to discriminate between arrays of three and six circles and a 10-year-old knowing that an array of three circles can, if one wished, be named with the semantic concept "3," which is symbolic of a magnitude less than the concept "4." Infants and adults would probably perceive a pile of playing cards composed of two full decks similarly, but only the adults know that they can conceptualize this event as one pile, two decks, or 104 cards.

Number is not an inherent property of a visual experience but one that is imposed on an event. An infant who discriminated between a cup three-fourths full of liquid and one one-fourth full does not possess the mathematical concept of fractions. Attributing number concepts to infants simply because they can discriminate among arrays containing different numbers of elements is analogous to attributing a number competence to pigeons who can be taught to peck at a key exactly four times (Machado & Rodrigues, 2007), to the sensory cells of the basilar membrane because they respond differently to varied sound frequencies, and to honeybees, whose behavior upon returning to the hive is affected by the distance between the hive and a flower bed.

If the claim that infants possess a concept of number, as contrasted with schemata for varied arrays of elements, is suspect the suggestion that young infants can "add" numbers must also be vulnerable to criticism. Wynn (1992, 1998) has written that young infants possess a mental mechanism devoted especially to number and can "add numerical quantities." The evidence for this bold declaration is that young infants looked longer at a stage containing one doll than at one with two dolls after seeing a sequence in which a human hand placed a second doll on a recently curtained stage that had contained one doll. But Wynn presented the infants with only two test events—either one or two dolls—and did not allow them a choice between one, two, or no dolls. Infants pay little attention to a stage containing no dolls, even though this is an unexpected events for infants who were administered the addition procedure (Cohen & Marks, 2002). This observation should raise some doubt regarding the claim that infants can add numbers. (See Bogartz et al., 1997, for another example of the importance of presenting more than two test events.) It is also relevant that the profile of attention shown by 5-month-old boys administered a partial replication of Wynn's procedure did not support her conclusion (Moore & Cocás, 2006).

More important, if infants were able to add numbers it should make no difference if they first saw an object on a stage and then saw a second object added behind a screen (Wynn's procedure) or initially saw no object on the stage but watched two objects being added sequentially to a stage located behind a screen. However, 8-month-olds exposed to this procedure did not behave as if they could add 1 + 1 (Uller, Carey, Huntley-Fenner, & Klatt, 1999). Furthermore, 3-year-olds behave as if they do not "know" an essential feature of the concept of number; namely, that the number of objects in an array remains the same despite a change in their spatial arrangement (Halford & Boyle, 1985). If preschool children do not possess this essential property of number, one can question the conclusion that the infants' representations of dolls moving on and off a stage share important features with the concept of number held by school-aged children. Notice that Wynn could have written that infants are able "to combine schemata for two visual events and hold the combined representation in a short term memory circuit." This conclusion would be less vulnerable to criticism, though still requiring proof, but it would not have made the headlines of our major newspapers. The predicate "add" has a specific meaning in mathematics and psychologists cannot treat that definition permissively. Differentiate and integrate have specific meanings in the calculus that are not the meanings understood by social psychologists.

It also remains a possibility that other processes operative in Wynn's procedure can explain the data. One such process involves the curvilinear discrepancy principle. The last event the infants saw in the procedure for addition was a single doll located at the edge of a curtained stage. The test event contained either two dolls or one doll in the middle of a stage with no curtain. One can argue that a single doll in the middle of a stage with no curtain is moderately discrepant from the last event the infant saw, whereas two dolls are more discrepant. Hence, looking at one doll would be predicted by the curvilinear principle.

Second, young infants often look at the more familiar event after seeing a sequence of new experiences. In the procedure for addition, three distinct events occurred before the test trial and on two of those occasions, the infants saw one doll and never saw two dolls. Thus, they might have looked longer at one rather than two dolls on the test trial because it was the more familiar event and not because they expected to see two dolls.

Furthermore, Cohen and Marks (2002) remind us that young infants usually look longer at the most recent, rather than at the more remote, event when exposed to a sequence of events the investigator assumes is perceived as connected because the young infants forgot the initial experience. Six to 7-month-olds first watched two different objects placed sequentially behind separate occluders on distinct places on a stage. Although they showed increased attention when the object that had been hidden last changed shape and color, they failed to do so when the object placed first behind the occluder underwent the same change (Kaldy & Leslie, 2005). The authors argued that infants less than 9 months, before the maturational transition associated with an enhancement of working memory, have difficulty retrieving the initial events of a sequence. The last event Wynn's infants saw in the addition procedure was a single doll. A similar interpretation can be applied to a later study with 9-month-olds watching films of five geometric forms moving or emerging from behind an occluder (McCrink & Wynn, 2004).

The young infant's ability to retain schemata for events experienced many seconds earlier and to retrieve them in the service of processing a new event, grows linearly during the 1st year (Fox, Kagan, & Weiskopf, 1979; Pelphrey et al., 2004; Ross-Sheehy et al., 2003). Most infants under 7 months have difficulty remembering the location of an object when there is only a 7- to 10-s delay between the hiding of the object and the time when the infant is allowed to reach for it because the connectivity of medial temporal lobe structures and prefrontal cortex is immature (Diamond, 1990). Infants younger than 8 months looking at pairs of faces consisting of one unchanging photo and a series of different unfamiliar faces required more than 60 s of exposure to the unchanging face before they looked consistently longer at the novel stimulus (Rose, Feldman, & Jankowski, 2002; see also Rose, Feldman, & Jankowski, 2001). Children less than 3 years old typically do not integrate relevant past experiences with their understanding of the present moment (Povinelli, Landry, Theall, Clark, & Castille, 1999).

In light of this evidence, it is surprising that Dehaene (2007) writes that psychological phenomena cannot be understood fully until they "have been related to all levels of brain organization" (p. 29) and contends that infants possess a concept of number localized in the intraparietal region of the brain. I assume Dehaene means that there should be a close correspondence between the behavioral and the biological evidence. If Dehaene is correct, the hypothesis localizing number in the intraparietal region of the infant's brain is suspect. First, the biological changes that occur in this region and in its connections after the 1st year could alter its functions. Second, native adult Chinese and English speakers show different patterns of brain activity while solving addition problems (Tang et al., 2006). Third, adults required to select the most likely answer (from two alternatives) to a series of multiplication problems showed the largest increase in blood flow in precentral and inferior gyri and displayed a decrease in blood flow to the intraparietal area (Ischebeck, Zamarian, Egger, Schocke, & Delazer, 2007). Fourth, a major function of the intraparietal area involves the perception of a spatial array of objects, mediated primarily by the right hemisphere (de Schonen & Mathivet, 1990; Hagler, Riecke, & Sereno, 2007). The fact that the intraparietal area is active when humans see a display of four circles is more properly described as a response to a spatially distributed array of discrete stimuli rather than a reaction to the recognition that there are four objects. Blood flow profiles to the intraparietal area were different to spatially distinct arrays of three objects (one array grouped two of the objects and the other did not; Xu & Chun, 2007). If this cortical area "counted" the number of objects, the blood flow patterns should have been similar to both arrays because they contained the same number. Furthermore, the cortical areas activated when adults were looking at arrays with varied numbers of objects were different from the areas active when they were reading numbers (Roux, Lubrano, Lauwers-Cances, Giussani, & Demonet, 2008). These facts, and others, are inconsistent with the hypothesis that the infant's behavior to visual displays varying in number of elements reflects a competence that is similar to the older child's concept of number when doing arithmetic calculations.

Inferring Physical Impossibility

The claim that infants "know" that solid objects cannot pass through a solid barrier has also generated strident debate and readers are referred to volume 1 of the journal *Infancy* for a set of important articles, both pro and con, on this issue (Aslin, 2000). Baillargeon (1994) believes that infants less than 8 months old know that one solid object cannot pass through another. In one experiment, 6.5-month-olds watched a screen move in a 180-degree arc on a stage, resembling the motion of a drawbridge. The examiner then placed a solid box behind the screen to inform the infant that the screen could no longer traverse the full 180 degrees because the box would obstruct its passage. On the test trial, some infants saw the screen traverse 157 degrees, an impossible event; others saw the screen move in an arc of 112 degrees, which is a possible event. The infants looked longer at the impossible event. But the last dynamic event the infant saw was the screen moving in an arc of 180 degrees; therefore, seeing it move 157 degrees was more similar to that event than the 111 degree motion and, in addition, involved more motion (Baillargeon, 1994). Infants who saw a screen move 112 or 180 degrees in a setting where both were possible events looked longer at the event with more motion, that is, the 180 degree movement (Rivera, Wakely, & Langer, 1999). Baillargeon suggested, however, that her evidence does not mean that infants respond to physical events "in terms of abstract underlying principles, but in terms of concrete categories corresponding to specific ways in which objects behave or interact" (p. 139). I interpret "concrete categories" to mean perceptually based schemata rather than semantically based conceptual representations of physical impossibility.

Schilling (2000) recognized that 4-month-olds might require a great deal of experience to create a firm schema for the complicated scene of a moving screen and an obstructing object on a stage. Hence, he gave one group of infants 7 exposures to the screen moving the full 180 degrees and gave another group 12 exposures to the same sequence. Schilling then placed an obstructing object on the stage (but one less salient than the brightly colored block used by Baillargeon, 1994) and had the screen rotate either 180 degrees or 112 degrees. The infants who had only seven exposures and therefore may not have habituated to the sequence, looked longer when the screen moved a full 180 degrees, the impossible event. However, the infants who had seen the 180 degree movement 12 times, and therefore were more likely to be bored, looked longer when the screen moved 112 degrees, the possible event that was discrepant from their more firmly formed schemata for the full 180 degree traverse. (See Baillargeon, 2000, for a rebuttal to Schilling, 2000.)

Schoner and Thelen (2006) have also questioned Baillargeon's (1994) conclusion that differential looking is a sufficient basis for assuming that infants know one object cannot pass through another. These scientists prefer the concepts of activation and inhibition fields, which vary continuously in strength and compete with each other. Schoner and Thelen assume that infants are prompted to look at a stimulus when the strength of the activation field motivating attention is stronger than the inhibition field motivating looking away. However, they agree with this author and others that the physical features of the stimulus and the timing and order of stimulus presentations can influence, in a serious way, the profiles of fixation time. These investigators have formalized the activation and inhibitory processes in a neural model that explains the evidence without reference to objects or physical impossibility.

In a second procedure designed to evaluate the infants' appreciation of physical impossibility, 7.5-month-olds were familiarized on one of two displays of a ramp, a track, and an agent object that could contact a target object. The infants in the potential contact group saw the agent and target separated by a half barrier; hence, the former theoretically could contact the latter. The infants in the "no-contact" group saw a full barrier between the ramp and the track; hence, the agent could not contact the target. The test display, which was similar for both groups, presented infants with the top half of the barrier but the bottom half covered with a curtain, making it appear as if there were a full barrier between ramp and track. The agent rolled down the ramp, passed behind the curtain, and seconds later the target moved on the track. Because the no-contact group looked longer at this event than the contact group, the authors concluded that these infants "knew" the agent could not strike the target and cause it to move (Kotovsky & Baillargeon, 2000). However, the test event was moderately discrepant from the familiarization experience for the no-contact infants because a curtain replaced the bottom half of the full barrier. The test stimulus was a more serious discrepancy for the "contact" group who had seen only a half-barrier during familiarization and no curtain. Thus, the curvilinear principle could account for these data.

It is also relevant that 6-month-olds did not show prolonged attention to a physically impossible movement sequence involving an agent and target object, and 10-month-olds, who saw both a possible and an impossible sequence without any prior familiarization, did not look longer at the physically impossible event (Oakes & Cohen, 1990). Moreover, 10-month-olds first familiarized on an impossible sequence behaved as if discrepancy was the primary determinant of attention for they looked longer at a possible sequence than at an impossible one (see Oakes, 1994, for a similar result).

Saxe and Carey (2006), who reviewed the research on the infant's ability to infer a necessary causal relation when an agent object contacts and dislocates a target object, concluded that infants become attentive when there is a violation of a perceptually based expectation of a target's movements, but they are not very attentive when the target changes color or state, suggesting that the target's motion has perceptual salience. Saxe and Carey suggested that perception of a "causal" relation between two events is not present at birth and, like the perception of the unity of an object, requires postnatal experiences. By the second half of the 1st year, infants have had a great deal of experience seeing hands throw or move objects and these experiences create schemata for hands as the origins of motion in another object. Hence, they look longer when a human hand is the agent that changes the object or location it approaches on a stage than when the agent is an oval green object (Shimizu & Johnson, 2004). Infants can learn perceptual expectations with surprising speed (Kochukhova & Gredeback, 2007). However, the important point is that the young infant's schematic representation of a sequence of physically plausible events is not a general competence but one that is limited to the classes of events that infants have encountered in the past. That is why the patterns of looking in 4-month-olds watching an object move in horizontal or diagonal trajectories, which included passing behind an occluder in a continuous or discontinuous trajectory, failed to support the hypothesis that young infants possess an understanding of the possible or impossible motions of objects (Bremner et al., 2007).

The expectation that an object will move when struck by another does not mean that infants understand why a movement did or did not occur (Schlottmann, 1999). Haith and Benson (1998) point out that when a moving object does not contact a target object and, following a brief delay, the latter moves (an impossible sequence), infants might be reacting to the difference in timing between the familiarization experience, in which there was simultaneity between the contact of agent with target and the subsequent motion of the target and the test event in which simultaneity was missing. Most 4-year-olds are unable to report the reasons why an impossible event cannot occur (Shtulman & Carey, 2007). Older children have learned that morning light always follows the dark of night and would be surprised if there were no light when they woke at 8 a.m., but they do not believe that the latter was the cause of the former.

Preschool children who have seen restaurant scenarios on television are likely to have a perceptual schema for a sequence in which an adult writes down on paper what sitting adults say and, minutes later, delivers some food to the latter. Hence, they would be surprised if the waiter brought the customer a toy car. But that fact does not mean that the child "knows" anything about the chain of events that transpired between the perception of the adult writing on a pad and the later delivery of the food. Many adults have a perceptual schema for hitting the button on their computer and seeing, seconds later, the word *Welcome* on the screen. But they do not possess a conceptual representation of the causal sequences that occurred between the two events. Rats can acquire an expectation of receiving food when they strike a lever and might show behavioral signs of surprise if the food did not appear, but few scientists would claim that the animals possessed a representation of the causal relation between their actions and the delivery of a few pellets. There is a great deal of evidence to suggest that the first signs of the ability to infer the invisible events that might mediate an outcome do not emerge until the middle of the 2nd year (Kagan, 1981). Causality, as Hume understood, is a symbolic invention and is not contained in a perceptual schema.

It is important to appreciate that 2-year-olds, who have some language and an improved working memory, do not infer the correct location of a ball that rolls down an incline to strike a tall, clearly visible occluder. In this study, the examiner placed the occluder behind one of the four doors over a series of trials and asked the child to guess which door should be opened in order to retrieve the ball. Two-year-olds could not solve this problem; that is, they were unable to infer the correct door to open from their perception of the location of the occluder (Berthier, De Bois, Poirier, Novak, & Clifton, 2000). However, 2-year-olds do show increased attention if a puppet, who had been successfully retrieving a ball from the correct door, failed to reach for the object at the door representing the correct location (Mash, Novak, Berthier, & Keen, 2006). Hood, Cole-Davies, and Dias (2003) performed a similar experiment with a similar result. Two-and-a-half-year-olds stared for several seconds when a toy that had rolled down a ramp and struck an occluder suddenly appeared on the other side of the occluder—a physically impossible event—but failed to search correctly for the toy at the place where the occluder rested.

These results affirm the earlier suggestion that the predicate "know" has two different meanings in these varied experimental contexts. One refers to prolonged attention following the violation of a perceptual expectation; the second refers to a conceptually richer representation that includes an inference representing a necessary relation between the location of an occluder and the location of an object. Because the 2-year-olds did not reach correctly for the object when they could see where the occluder rested, investigators may be going beyond the evidence when they conclude that infants less than 1 year "know" that objects cannot pass through a solid barrier and imply that "know" has the same meaning for infants that it has for older children.

Object Permanence

Eight-month-olds administered the Piagetian A not B task often stare briefly at the correct location of a hidden object but reach incorrectly (Ahmed &

Ruffman, 1998; Hofstadter & Reznick, 1996). Some psychologists have used the staring response as evidence that infants "know" the correct location of the object. But neuroscientists have learned that the dorsal circuit, which influences the frontal eye fields in the left hemisphere and mediates the location of an event (the "where" circuit), matures earlier than the ventral circuit that mediates the representation of an object-qua-object (the "what" circuit; Campana, Cowey, Casco, Oudsen, & Walsh, 2007; Rao, Zhou, Zhou, Fan, & Chen, 2003). Knowing the location where an event just occurred is not equivalent to knowing that a particular object, seen hidden seconds earlier, is resting in that location. I will orient in the direction of a fast-moving object detected in my peripheral vision without having any knowledge of the features or category of the object. Hence, looking at the correct location where a toy had been hidden may not represent a point on a continuum of object permanence but rather a totally different cognitive process (Ruffman, Slade, Carlos-Sandino, & Fletcher, 2005). Some dog breeds will run hundreds of yards to retrieve a stick thrown by their owner even though they cannot see the object they are racing to retrieve. It is not obvious that the dog knows that the stick is a permanent object that cannot disappear.

In summary, the increased fixation time observed after violation of a perceptual expectancy is influenced by at least six different factors, depending on experimental conditions: (a) the event seen last in a sequence and the timing of the events in the sequence, (b) the more familiar event in a complex sequence, (c) the more unfamiliar of two test events, (d) an unexpected event, (e) the moderately discrepant event when more than two unfamiliar events are presented, and always (f) the salient physical features of the event. No investigation, to my knowledge, has adequately controlled for all six factors. Hence, it is reasonable to question the claim that 4-month-olds have a concept of number, can add, and "know" why an object cannot pass through a solid substance. The same skepticism applies to the claim that infants know that an illustration of a cube, resembling an Escher drawing, is physically impossible to construct. Careful examination of the details in the possible and impossible cubes reveals subtle differences in their physical features (Shuwari, Albert, & Johnson, 2007).

Including Other Measures

Because all conclusions are constrained by their source of evidence, it is relevant that measures other than total fixation time, including operant responding and changes in heart rate, skin conductance, and vagal tone, occasionally invite different inferences about an infant's knowledge (Bornstein & Suess, 2000; Shields & Rovee-Collier, 1992). The various waveforms in the event-related potential (ERP) also lead to different conclusions regarding the state of an infant following violation of an expectation. For example, 6-month-olds who saw, in a random series, a familiar and unfamiliar face, and a familiar and unfamiliar object showed a larger negative waveform at about 400 ms (called Nc) to the familiar events, but a larger positive slow waveform at about 800 ms to the unfamiliar stimuli (de Haan & Nelson, 1999).

Infants between 4.5 and 7.5 months were first familiarized on two geometric forms (A and B) and later saw form A on 60% of the trials, form B on 20% of the trials, and a third novel form (C) on 20% of the trials. The Nc waveform was largest to the novel form C. But had attention been coded in the usual fashion many infants would have looked longest at form B, the moderately discrepant event (Reynolds & Richards, 2005). Infants familiarized on trios of objects that varied in shape, color, and orientation over trials failed to show an increase in attention to the unexpected appearance of two objects but did display a clear change in an ERP waveform to two objects (Arriaga, 2001). Finally, there was no correlation in 7-month-olds between the difference in fixation time to faces displaying a fearful compared with a happy expression (larger values to the fear face) and the differences in the magnitudes of the Nc or P400 waveforms to these two facial expressions, despite larger values to the fear face (Leppanen, Moulson, Vogel-Farley, & Nelson, 2007; see de Haan & Nelson, 1997; Slater, Morison, & Somers, 1988; Webb, Long, & Nelson, 2005, for additional support of the claim that different conclusions follow from data on looking time compared with ERP evidence).

Vocalization, smiling, and changes in facial expression can also reflect psychological states provoked by violations of expectations that looking times fail to reveal. For example, infants looked longer at a speaker baffle playing the babbling of an unfamiliar infant than at one playing their own babbling, but the infants vocalized more to the latter event, even though they did not usually hear tape recordings of infant babbling and, therefore, both events were unexpected. Four-month-olds playing peek-a-boo with an adult who suddenly displayed a sad, rather than a smiling, face showed no change in fixation time but displayed a change in facial expression (Montague & Walker-Andrews, 2001). On occasion, smiling reflects recognition of a discrepant event when fixation time does not. Three-month-olds showed equivalent fixation times to pictures of a normal face and a face with a single cyclopean eye in the middle of the forehead but only smiled to the regular face (Lewis, 1967).

Prolonged attention to an event appears to reflect the state created by an event that engages a schema but is not assimilated immediately. Babbling or smiling reflects the state created by a discrepant event that is easier to match to a schema and is assimilated more rapidly (Kagan et al., 1978). Moreover, the combination of fixation time and smiling, coded simultaneously, was the best predictor of how many exposures to a familiarization stimulus were required before 5- to 9-month-old infants looked reliably longer at the novel member of a pair of events (Rose, Futterweit, & Jankowski, 1999).

An elegant experiment with monkeys revealed the inferential error that is possible when investigators use only one measure to evaluate a cognitive process or structure. Four rhesus monkeys were trained over many days to make a simple motor response whenever a particular human face appeared on a screen. After the animals had learned to discriminate this face from many others, the psychologists altered the features of this target face in order to discover the essential features of the monkeys' schema for the familiar face. As expected, the animals were less likely to make the motor response (indicating recognition) when the size or location of the eyes was altered than when there were alterations in the nose or chin, suggesting that the size and placement of the eyes were the most essential features of the monkeys' schema for the target face. But, surprisingly, some animals showed no change in behavior to a face with no eyes at all and behaved as if this stimulus were the target face (Keating & Keating, 1993).

This unexpected result invites one reasonable and one unreasonable explanation. The latter implies that monkeys perceive faces in a manner that is qualitatively different from humans; no child would treat a face with no eyes as equivalent to one with eyes. If this explanation were correct, a great deal of theory would have to be rewritten. Hence, the more reasonable interpretation is that the monkey's behavior is ambiguous in meaning and can be evoked by more than one psychological state. I suspect this is the proper inference and, therefore, urge more caution among scientists who make claims about what infants know that are based on only one measure, whether that measure is fixation time, instrumental kicking, ERPs, autonomic response, smiling, or babbling. There is, at present, no behavioral or biological reaction that is restricted to a single set of conditions and demands only one inference. Readers are reminded of the earlier commitment by behaviorists to the concept of the strength of a habit. However, this a priori concept lost its attractiveness when research revealed that the three measures of habit strength in a hungry rat—speed of running down an ally to a goal box, making the correct turn in a maze, and amount of food consumed after arriving at the goal—were not highly correlated.

Influence of Brain Maturation

The lawful alterations in brain structure, chemistry, and physiology over the first decade raise the possibility of qualitative changes in the psychological structures that represent objects, number, or causality in infants compared with older children (Lenroot & Giedd, 2007). If the brains of 6-month-olds were essentially similar to those of 16-year-olds, the nativist argument would be more attractive, but this is not the case (Almli, Rivkin, McKinstry, & the Brain Development Cooperative Group, 2007). For example, during the second half of the 1st year, there is a major spurt of differentiation in both pyramidal and inhibitory interneurons in the prefrontal cortex, a broader distribution of the double bouquet interneurons in Layer 3, an increase in the number of spines and extra large excrescences on the proximal dendrites of pyramidal cells in the CA3 region of Ammon's horn of the hippocampus, and increased glucose uptake in the lateral, frontal, and dorsolateral prefrontal cortex (Chugani, 1994; Huttenlocher, 1974, 1990; Kostovic, 1987, 1990; Mrzljak, Uylings, Van Eden, & Judas, 1990). Furthermore, the connections between the amygdala–temporal lobe and the prefrontal cortex are strengthened during the second half of the 1st year, in part because axons within the capsula interna develop myelin (Brody, Kinney, Kloman, & Gillis, 1987; Chrousos & Gold, 1992). It cannot be a coincidence that these structural changes occur at the same time psychologists observe a significant improvement in working memory (Reznick, Morrow, Goldman, & Snyder, 2004). These biological changes permit 1-year-olds to hold representations of experience in an active state for a longer time and to retrieve representations of events that occurred 10–30 s earlier (Kagan & Herschkowitz, 2005; Seress & Mrzljak, 1992; Witter, 1993).

Not surprisingly, these structural changes are accompanied by changes in power in the electroencephalogram (EEG) and ERPs. For example, a working memory task (Piaget's A not B problem) administered to 8-month-olds evoked changes in power across the entire cortex, whereas a working memory task administered to 4-year-olds was associated with more localized changes in power restricted to medial frontal sites and increased coherence between frontal and temporal–occipital sites (Bell & Wolfe, 2007). The various waveforms of the ERP also undergo a major change between 3 and 24 months (Jing & Benasich, 2006; Richards, 2003; Scott, 2006; Webb et al., 2005), and the amplitudes of the P2/N2 waveform to click sounds increase in a linear fashion between 1 and 12 months (Ohlrich & Barnet, 1972). In addition, the proportion of quiet sleep associated with sleep spindles (12–14 Hz) increases in precentral sites at 9–10 months (Tanguay, Ornitz, Kaplan, & Bozzo, 1975).

The brain continues to mature after the first birthday. Blood flow is greater to the right than to the left hemisphere during the opening 3 years but shifts to more predominant flow to the left hemisphere at about 4 years of age (Chiron et al., 1997). Synaptic density in the primary auditory cortex does not reach adult values until puberty; dendrites on neurons of the dentate nucleus of the cerebellum undergo intensive lengthening and extension, and the basal ganglia continue to develop during the 2nd year. Dopamine receptors in the caudate nucleus (especially D1 and D2 receptors) do not reach maximal density until the middle of the 2nd year and the neurons of Cortical Layer 3, which comprise the cells of the corpus callosum, increase in size during this interval (Huttenlocher & Dabholkar, 1997).

The concentration of the enzyme involved in the synthesis of acetylcholine increases sharply during the 2nd year when pyramidal neurons begin to express acetylcholinesterase in cell bodies and fibrillary networks (Black, 1991; Blokland, 1996). The innervation of Layer 3 pyramidal neurons by acetylcholinesterase and the growth of Layer 3 neurons probably contribute to the seminal advances of the 2nd year, especially the emergence of language, a moral sense, and self-awareness (Kagan, 1981; Mesulam & Geula, 1991). Finally, the human brain does not attain 90% of its adult weight until 7 or 8 years of age (Giedd et al., 1999), the density of D1 receptors in the basal ganglia does not attain peak density until 4–8 years (Seeman et al., 1987), the degree of coherence of various EEG frequency bands between frontal and occipital or temporal areas increases after the third birthday, and the latency to the peak value of the N170 waveform to faces decreases by as much as 10% between 5 and 7 years of age (Taylor, McCarthy, Saliba, & Degiovanni, 1999).

In sum, between the first and the eighth birthdays, the human brain attains peak synaptic density, maximal glucose uptake in the prefrontal cortex, peak concentrations of dopamine and norepinephrine receptors, myelination of long cortical tracts, coherence in the EEG, and maximal differentiation of dendrites in the hippocampus. It seems reasonable to suggest that these changes are accompanied by qualitative transformations of the cognitive functions of infants.

Final Summary

The most enthusiastic advocates of a return to nativism do not insist that the infants' concepts for number, physical impossibility, causality, or the permanence of objects are essentially the same as those of adolescents. The central suggestion in this article is that the perceptual representations that comprise schemata are to be distinguished from the conceptual structures involving semantic forms and inferences about events that were not experienced. Infants turn away from dissonant musical chords but do not possess the musician's concept of dissonance; infants who grasp an object placed in their palm do not understand the concept "mine." Spelke (1994), an influential advocate and sophisticated experimental psychologist, reminds us that there is no consensus as to when particular knowledge structures appear, how they are altered with growth and experience, what role the infants' representations play in the later development of thought and reason, and adds, wisely, that the infants' cognitive functioning has to differ from that of adults because "distinct knowledge systems are less interconnected" during the 1st year of life (Spelke, Vishton, & Von Hofsten, 1995, p. 177).

Hence, it is appropriate to ask why the notion that young infants possess conceptual talents resembling those of older children became attractive to a select cohort of developmental psychologists in such a short time. One speculative explanation is that there is a competition for dominance between the language networks of biology and those of the social sciences. The dramatic advances in molecular biology and neuroscience, which are to be celebrated, have led most biologists, but a smaller number of social scientists, to assume that the vocabulary of the brain sciences will prove to be more fruitful than the psychological concepts that emerge from brain activity and depend on the products of experience. This article has suggested that a construct purporting to explain a psychological competence as a process inherent in the infant's brain organization need not have the same meaning as one that explains a process requiring extensive experience. However, some younger scientists have become attracted to the language networks of biology and neuroscience because these domains have become challenging areas of inquiry that provide generous research funds to those who study the relation between biological and psychological outcomes. Scientists who study the relations between experience and psychological phenomena find it more difficult to gain support of their work. Even the most committed scientists are susceptible to the schedule of rewards. I suspect that some social scientists feel that the importance of their research is enhanced if their ideas and data can be classified as belonging to biology rather than psychology. The idea that mathematical and physical principles are represented in the brains of 6-month-olds seems to be a biological, rather than a psychological, fact.

A second reason is that psychologists have tried for many decades to specify the experiences that a maturing brain, with its biologically prepared sensory and motor skills, exploits as children develop new competences. Because investigators have not been as successful as they hoped, their frustration is reduced a little by declaring that the kernel of the conceptual ability was already present in the first half of the 1st year. If this suggestion were true, the difficult task of discovering the differences between the earlier and the later forms, and understanding how those differences were created, would recede into the shadows. If 4-month-olds already appreciate that some physical events are impossible, the fact that 3-year-olds possess a kernel of this principle becomes less of a mystery. However, this position is surprising in light of what biologists have learned about embryological development. The 3-week-old embryo possesses neither a convoluted cerebral cortex nor the structural changes that the secretion of steroid hormones will induce at a later time.

The next cohort of developmental psychologists interested in these questions should gather additional sources of evidence besides fixation time; these include ERPs, skin conductance, muscle tension, blood flow, changes in heart rate, smiling, and vocalizing. Second, they should try to discover, with cross-sectional or longitudinal designs, the experiences necessary for actualization of the competences in older children that might assume a simpler form in infants (Campos & Bertenthal, 1989). This evidence could illuminate the nature of the changes that occur between infancy and later childhood and clarify the degree of continuity or discontinuity between them.

If 2-year-olds cannot add the cardinal numbers 2 and 4 and do not know which door to open to retrieve a ball resting behind a visible occluder, it is possible that the apparently similar "knowledge" infants are presumed to possess is qualitatively different from that of the school-aged child. I agree with Aslin's (2000) suggestion that the nativist position remains viable until additional evidence clearly refutes or fails to reject this view. Additional research is required before psychologists can be certain of the degree of similarity in mechanism between the cognitive functions of infants and older children. My skepticism rests on the observation that it is rare, in any scientific domain, for a single measure to reveal all that is necessary to understand a phenomenon. Hence, it is reasonable to question the assumption that an increase in fixation time of a few seconds to an event that violates an acquired perceptual expectation, without any other evidence, is a sufficient basis for inferring that infants under 6 months possess a conceptual representation of number, physical impossibility, causality, or the permanence of objects that shares important features with the competences awarded the same name in older children.

References

Ahmed, A., & Ruffman, T. (1998). Why do infants make A not B errors in a search task, yet show memory for the location of hidden objects in a nonsearch task? *Developmental Psychology, 34*, 441–453.

Almli, C. R., Rivkin, M. J., McKinstry, R. C., & the Brain Development Cooperative Group. (2007). The NIH MRI Study of normal brain development: Newborns, infants, toddlers, and preschoolers. *Neuroimage, 35*, 308–325.

Arriaga, R. (2001). *Behavioral and electrophysiological correlates of numerosity discrimination in eight-month-old infants*. Unpublished doctoral dissertation, Harvard University, Cambridge, MA.

Aslin, R. N. (2000). Why take the cog out of infant cognition? *Infancy, 1*, 463–470.

Aslin, R. N. (2007). What's in a look? *Developmental Science, 10*, 48–53.

Baillargeon, R. (1994). How do infants learn about the physical world? *Current Directions in Psychological Science, 3*, 133–140.

Baillargeon, R. (2000). Reply to Bogartz, Shinskey, and Schilling; Schilling; and Cashon and Cohen. *Infancy, 1*, 447–462.

Bar, M., & Neta, M. (2007). Visual elements of subjective preference modulate amygdala activation. *Neuropsychologia, 45*, 2191–2200.

Beeman, M., & Chiarello, C. (1998). *Right hemisphere language comprehension*. Mahwah, NJ: Erlbaum.

Bell, M. A., & Wolfe, C. D. (2007). Changes in brain functioning from infancy to early childhood. *Developmental Neuropsychology, 31*, 21–38.

Berthier, N. E., De Bois, S., Poirier, C. R., Novak, M. A., & Clifton, R. K. (2000). Where's the ball? *Developmental Psychology, 36*, 394–401.

Black, I. B. (1991). *Information and the brain*. Cambridge, MA: MIT Press.

Blokland, A. (1996). Acetylcholine: A neurotransmitter for learning and memory. *Brain Research Review, 21*, 285–300.

Bogartz, R. S., Shinskey, J., & Speaker, C. J. (1997). Interpreting infant looking: The event set x event set design. *Developmental Psychology, 33*, 408–422.

Bogartz, R. S., Shinskey, J. L., & Schilling, T. H. (2000). Object permanence in five-and-one-half-month-old infants? *Infancy, 1*, 403–428.

Bornstein, M., & Suess, P. E. (2000). Physiological self-regulation and information processing in infancy: Cardiac vagal tone and habituation. *Child Development, 71*, 273–287.

Bremner, J. G., Johnson, S. P., Slater, A., Mason, U., Cheshire, A., & Spring, J. (2007). Conditions for young infants' failure to perceive trajectory continuity. *Developmental Science, 10*, 613–624.

Brennan, W., Ames, E. W., & Moore, R. W. (1966). Age differences in infants' attention to patterns of different complexities. *Science, 151*, 354–356.

Brody, B. A., Kinney, H., Kloman, A., & Gillis, F. H. (1987). Sequence of central nervous system myelination in human infancy. 1. An autopsy study of myelination. *Journal of Neuropathology and Experimental Neurology, 46*, 283–301.

Campana, G., Cowey, A., Casco, C., Oudsen, I., & Walsh, V. (2007). Left frontal eye field remembers "where" but not "what." *Neuropsychologia, 45*, 2340–2345.

Campos, J. J., & Bertenthal, B. I. (1989). Locomotion and psychological development in infancy. In F. Morrison, C. Lord, & D. Keating (Eds.), *Applied developmental psychology* (Vol. 2, pp. 230–258). New York: Academic Press.

Cashon, C. H., & Cohen, L. B. (2000). Eight-month–old infants' perception of possible and impossible events. *Infancy, 1*, 429–446.

Chiron, C., Jambaque, I., Nabbout, R., Lounes, R., Syrota, A., & Dulac, O. (1997). The right brain hemisphere is dominant in human infants. *Brain, 120*, 1057–1065.

Chrousos, G. P., & Gold, P. W. (1992). The concepts of stress and stress system disorders. *Journal of the American Medical Association, 267*, 1244–1252.

Chugani, H. T. (1994). Development of regional brain glucose metabolism. In G. Dawson & K. Fischer (Eds.), *Human behavior and the developing brain* (pp. 153–175). New York: Guilford.

Cohen, L. B., & Cashon, C. H. (2006). Infant cognition. In W. Damon & R. M. Lerner (Series Eds.) & D. Kuhn & R. Siegler (Vol. Eds.), *Handbook of child psychology* (Vol. 2, 6th ed., pp. 214–251). New York: Wiley.

Cohen, L. B., & Marks, K. S. (2002). How infants process addition and subtraction events. *Developmental Science, 5*, 186–201.

Courage, M. L., Reynolds, G. D., & Richards, J. E. (2006). Infants' attention to patterned stimuli: Developmental change from 3 to 12 months of age. *Child Development, 77*, 680–695.

Cycowicz, Y. M., & Friedman, D. (2007). Visual novel stimuli in an ERP novelty oddball paradigm. *Psychophysiology, 44*, 11–29.

de Haan, M., & Nelson, C. A. (1997). Recognition of the mother's face by six-month old infants. *Child Development, 68*, 187–210.

de Haan, M., & Nelson, C. A. (1999). Brain activity differentiates face and object processing in six-month-old infants. *Developmental Psychology, 35*, 1113–1121.

Dehaene, S. (1997). *The number sense.* New York: Oxford University Press.

Dehaene, S. (2007). A few steps toward a science of mental life. *Mind, Brain, and Education, 1*, 28–47.

Deloche, G., Seron, X., Scius, G., & Segui, J. (1987). Right hemisphere language processing. *Brain and Language, 30*, 197–205.

de Schonen, S., & Mathivet, E. (1990). Hemisphere asymmetry in a face discrimination task in infants. *Child Development, 61*, 1192–1205.

Diamond, A. (1990). Developmental and neural bases of memory functions as indexed by the AB and delayed response tasks in human infants and infant monkeys. *Annals of the New York Academy of Sciences, 608*, 267–309.

Doupe, A. J., & Kuhl, P. K. (1999). Birdsong and human speech. *Annual Review of Neuroscience, 22*, 567–631.

Ellis, H. (1900). The analysis of the sexual impulse. *Alienist and the Neurologist, 21*, 247–262.

Fantz, R. L., & Fagan, J. F. (1975). Visual attention to size and number of pattern details by term and pre-term infants during the first six months. *Child Development, 46*, 3–18.

Feigenson, L., & Carey, S. (2005). On the limits of infants' quantification of small object arrays. *Cognition, 97*, 295–313.

Feigenson, L., Carey, S., & Spelke, E. (2002). Infants' discrimination of number vs. continuous extent. *Cognitive Psychology, 44*, 33–66.

Fenwick, K. D., & Morrongiello, D. A. (1998). Spatial colocation and infants' learning of auditory-visual associations. *Infant Behavior and Development, 21*, 745–760.

Fiske, J. (1883). *The meaning of infancy.* Boston: Houghton-Mifflin.

Fox, N., Kagan, J., & Weiskopf, S. (1979). The growth of memory during infancy. *Genetic Psychology Monographs, 99*, 91–130.

Frick, M., & Colombo, J. (1996). Individual differences in infant visual attention: Recognition of degraded visual forms by four-month olds. *Child Development, 67*, 188–204.

Giedd, J. N., Blumenthal, J., Jeffries, N. O., Rajapakse, J. C., Vaituzis, A. C., & Liu, H. (1999). Development of the human corpus callosum during childhood and adolescence: A longitudinal MRI study. *Progress in Neuropsychopharmacology and Biological Psychiatry, 23*, 571–588.

Hagler, D. J., Riecke, L., & Sereno, M. I. (2007). Parietal and superior frontal visuospatial maps activated by pointing and saccades. *Neuroimage, 35*, 1562–1577.

Haith, M. M. (1998). Who put the cog in infant cognition? *Infant Behavior and Development, 21*, 167–179.

Haith, M. M., & Benson, J. B. (1998). Infant cognition. In W. Damon (Series Ed.) &D. Kuhn & R. S. Siegler (Vol. Eds.), *Handbook of child psychology* (Vol. 2, 5th ed., pp. 199–254). New York: Wiley.

Halford, G. S., & Boyle, F. M. (1985). Do young children understand conservation of number? *Child Development, 56*, 165–176.

Hamburger, V. (1975). Changing concepts in developmental neurobiology. *Perspectives in Biology and Medicine, 20*, 162–178.

Hofstadter, M., & Reznick, J. S. (1996). Response modality affects human infant delayed- response performance. *Child Development, 67*, 646–658.

Hood, B., Cole-Davies, V., & Dias, M. (2003). Looking and search measures of object knowledge in preschool children. *Developmental Psychology, 38*, 61–70.

Hopkins, J. R., Zelazo, P. R., Jacobson, S. W., & Kagan, J. (1976). Infant reactivity to stimulus schema discrepancy. *Genetic Psychology Monographs, 93*, 17–62.

Houston, D. M., & Jusczyk, P. W. (2003). Infants' long term memory for the sound patterns of words and voices. *Journal of Experimental Psychology, 19*, 1143–1154.

Huttenlocher, P. R. (1974). Dendritic development in neocortex of children with mental defect and infantile spasms. *Neurology, 24*, 203–210.

Huttenlocher, P. R. (1990). Morphometric study of human cerebral cortex development. *Neuropsychologia, 28*, 517–527.

Huttenlocher, P. R., & Dabholkar, A. S. (1997). Regional differences in synaptogenesis in human cerebral cortex. *Comparative Neurology, 387*, 167–178.

Ischebeck, A., Zamarian, L., Egger, K., Schocke, M., & Delazer, M. (2007). Imaging early practice effects in arithmetic. *Neuroimage, 36*, 993–1003.

Ivry, R. B., & Lebby, P. C. (1998). The neurology of consonant perception. In M. Beeman & C. Chiarello (Eds.), *Right hemisphere language comprehension* (pp. 3–25). Mahwah, NJ: Erlbaum.

Jacobson, S. W. (1979). Matching behavior in the young infant. *Child Development, 50*, 425–430.

Jing, H., & Benasich, A. A. (2006). Brain responses to tonal changes in the first two years of life. *Brain Development, 28*, 247–256.

Johnson, S. P., Bremner, J. G., Slater, A., Mason, U., Foster, K., & Cheshire, A. (2003). Infants' perception of object trajectories. *Child Development, 74*, 94–108.

Jones, S. S. (1996). Imitation or exploration? *Child Development, 62*, 1952–1969.

Joseph, J. E., Powell, D. K., Andersen, A. H., Bhatt, R. S., Dunlap, M. K., Foldes, S. T., et al. (2006). fMRI in alert behaving monkeys: An adaptation of the human infant familiarization novelty preference procedure. *Journal of Neuroscience Methods, 157*, 10–24.

Kagan, J. (1981). *The second year.* Cambridge, MA: Harvard University Press.

Kagan, J. (2002). *Surprise, uncertainty, and mental structures.* Cambridge, MA: Harvard University Press.

Kagan, J., & Herschkowitz, N. (2005). *A young mind in a growing brain.* Mahwah, NJ: Erlbaum.

Kagan, J., Kearsley, R. B., & Zelazo, P. R. (1978). *Infancy.* Cambridge, MA: Harvard University Press.

Kagan, J., Klein, R. V., Haith, M. M., & Morrison, F. J. (1973). Memory and meaning in two cultures. *Child Development, 44*, 21–23.

Kagan, J., Linn, S., Mount, R., & Reznick, J. S. (1979). Asymmetry of inference and the dishabituation paradigm. *Canadian Journal of Psychology, 33*, 288–304.

Kaldy, Z., & Leslie, A. (2005). A memory span of one? Object identification in 6.5 month old infants. *Cognition, 97*, 153–177.

Keating, C., & Keating, E. G. (1993). Monkeys and mug shots. *Journal of Comparative Psychology, 107*, 131–139.

Kochukhova, O., & Gredeback, G. (2007). Learning about occlusion: Initial assumptions and rapid adjustments. *Cognition, 105*, 26–46.

Kostovic, I. (1987). Late postnatal development of acetylcholinesterase-reactive innervation of layer III pyramidal neurons in human prefrontal cortex. *Neuroscience*, 22, S228.

Kostovic, I. (1990). Structural and histochemical reorganization of the human prefrontal cortex during perinatal and postnatal life. *Progress in Brain Research, 85*, 223–240.

Kotovsky, L., & Baillargeon, R. (2000). Reasoning about collisions involving inert objects in 7.5 month old infants. *Developmental Science, 3*, 344–359.

Le Corre, M., & Carey, S. (2007). One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition, 105*, 395–438.

Lenroot, R. K., & Giedd, J. N. (2007). The structural development of the human brain as measured longitudinally with magnetic resonance imaging. In D. Coch, K. W. Fischer, & G. Dawson (Eds.), *Human behavior, learning, and the developing brain* (pp. 50–73). New York: Guilford.

Leppanen, J. M., Moulson, M. C., Vogel-Farley, V. K., & Nelson, C. A. (2007). An ERP study of emotional face processing in the adult and infant brain. *Child Development, 78*, 232–245.

Lewis, M. (1967). The meaning of a response or why researchers in infant behavior should be oriental metaphysicians. *Merrill-Palmer Quarterly, 13*, 7–18.

Lorenz, K. Z. (1958). The evolution of behavior. *Scientific American, 199*, 67–78.

Machado, A., & Rodrigues, P. (2007). The differentiation of response numerosities in the pigeon. *Journal of Experimental Analysis of Behavior, 88*, 153–178.

Mandler, J. M. (2004). *The foundations of mind: Origins of conceptual thought.* New York: Oxford University Press.

Marcus, G. F., Vijayan, S., Rao, S., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science, 283*, 77–80.

Mash, C., Novak, E., Berthier, N. E., & Keen, R. (2006). What do two-year-olds understand about hidden-objects? *Developmental Psychology, 42*, 263–271.

McCall, R. B., Eichorn, D. H., & Hogarty, P. S. (1977). Transitions in early mental development. *Monographs of the Society for Research in Child Development, 42*(Serial No. 171).

McCall, R. B., & McGee, P. (1976). *The discrepancy hypothesis of attention and affect in infants*. Unpublished manuscript.

McCrink, K., & Wynn, K. (2004). Large-number addition and subtraction by 9-month old infants. *Psychological Science, 15*, 776–781.

Meltzoff, A. N., & Moore, M. K. (1977). Imitation of facial and manual gestures by the human neomate. *Science, 198*, 75–78.

Mesulam, M. M., & Geula, C. (1991). Acetylcholinesterase rich neurons in the human cerebral cortex: Cytoarchitectonic and ontogenetic patterns. *Journal of Comparative Neurology, 306*, 193–220.

Mix, K. S., Huttenlocher, J., & Levine, S. C. (2002). Multiple cues for quantification in infancy. *Psychological Bulletin, 128*, 278–294.

Montague, D. P. F., & Walker-Andrews, A. S. (2001). Peekaboo: A new look at infants' perception of emotion expressions. *Developmental Psychology, 37*, 826–836.

Moore, D. S., & Cocas, L. A. (2006). Perception precedes computation. *Developmental Psychology, 42*, 666–678.

Moore, M. K , & Meltzoff, A. N. (2004). Object permanence after a 24-hr delay and leaving the locale of disappearance: The role of memory, space, and identity. *Developmental Psychology, 40*, 606–620.

Mrzljak, L., Uylings, H. B., Van Eden, C. G., & Judas, M. (1990). Neuronal development in human prefrontal cortex in prenatal and postnatal stages. *Progress in Brain Research, 85*, 185–222.

Nakagawa, A., Sukigara, M., & Benga, A. (2003). The temporal relationship between reduction of early imitative responses in the development of attention mechanisms. *BMC Neuroscience, 4*, 33–50.

Oakes, L. (1994). Development of infants' use of continuity cues and their perception of causality. *Developmental Psychology, 30*, 869–879.

Oakes, L., & Cohen, L. (1990). Infant perception of a causal event. *Cognitive Development, 5*, 183–207.

Ohlrich, E. S., & Barnet, A. B. (1972). Auditory-evoked responses during the first year of life. *Electroencephalography and Clinical Neurophysiology, 32*, 161–169.

Pelphrey, K. A., Reznick, J. S., Goldman, B. D., Sasson, N., Morrow, J., Donahue, A., et al. (2004). Development of visuospatial short-term memory in the second half of the first year. *Developmental Psychology, 40*, 836–851.

Plunkett, K., Hu, J. F., & Cohen, L. B. (2008). Labels can override perceptual categories in early infancy. *Cognition, 106*, 665–681.

Povinelli, D. J., Landry, A. M., Theall, L. A., Clark, B. R., & Castille, C. M. (1999). Development of children's understanding that the recent past is causally bound to the present. *Developmental Psychology, 35*, 1426–1439.

Rakic, P. (1995). Radial versus tangential migration of neural clones in the developing cerebral cortex. *Proceedings of the National Academy of Sciences, 92*, 11323–11327.

Rao, H., Zhou, T., Zhou, Y., Fan, S., & Chen, L. (2003). Spatiotemporal activation of the two visual pathways in form discrimination and spatial location. *Human Brain Mapping, 18*, 79–89.

Reynolds, G. D., & Richards, J. E. (2005). Familiarization, attention, and recognition memory in infancy. *Developmental Psychology, 41*, 598–615.

Reznick, J. S., Morrow, J. D., Goldman, B. D., & Snyder, J. (2004). The onset of working memory in infants. *Infancy, 6*, 145–154.

Richards, J. E. (2003). Attention affects the recognition of briefly presented visual stimuli in infants. *Developmental Science, 6*, 312–328.

Rivera, S. M., Wakely, A., & Langer, J. (1999). The drawbridge phenomenon: Representational reasoning or perceptual preference? *Developmental Psychology, 35*, 427–435.

Rochat, P., Striano, T., & Blatt, L. (2002). Differential effect of happy, neutral, and sad, still faces on two, four, and six-month-old infants. *Child Development, 11*, 289–303.

Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2001). Visual short-term memory in the first year of life: Capacity and recency effects. *Developmental Psychology, 37*, 539–549.

Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2002). Processing speed in the 1st year of life: A longitudinal study of pre-term and full-term infants. *Developmental Psychology, 38*, 895–902.

Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2003). Infant visual recognition memory: Independent contributions of speed and attention. *Developmental Psychology, 39*, 563–571.

Rose, S. A., Futterweit, C. R., & Jankowski, J. J. (1999). The relation of affect to attention and learning in infancy. *Child Development, 70*, 549–559.

Ross, S. M., & Dannemiller, J. L. (1999). Color contrast, luminance contrast, and competition in exogenous orienting in 3.5-month-old infants. *Infant Behavior and Development, 22*, 383–404.

Ross-Sheehy, S., Oakes, L. M., & Luck, S. J. (2003). The development of visual short-term memory capacity in infants. *Child Development, 74*, 1807–1822.

Roux, F. E., Lubrano, V., Lauwers-Cances, V., Giussani, C., & Demonet, J. F. (2008). Cortical areas involved in Arabic number reading. *Neurology, 70*, 210–217.

Ruffman, T., Slade, L., Carlos-Sandino, J., & Fletcher, A. (2005). Are A-not-B errors caused by a belief about object location? *Child Development, 76*, 122–136.

Saffran, J. R. (2001). Words in a sea of sounds. *Cognition, 81*, 149–169.

Saffran, J. R., Loman, M. M., & Robertson, R. R. W. (2000). Infant memory for musical experiences. *Cognition, 77*, 1315–1327.

Saffran, J. R., & Thiessen, E. D. (2003). Pattern induction by infant language learners. *Developmental Psychology, 39*, 484–494.

Saxe, R., & Carey, S. (2006). The perception of causality in infancy. *Acta Psychologica, 123*, 144–165.

Saxe, R., Tzelnic, T., & Carey, S. (2007). Knowing who dunnit: Infants identify the causal agent in an unseen causal interaction. *Developmental Psychology, 43*, 149–158.

Schilling, T. H. (2000). Infants' looking at possible and impossible screen rotations: The role of familiarization. *Infancy, 1,* 389 – 402.

Schlottmann, A. (1999). Seeing it happen and knowing how it works: How children understand the relation between perceptual causality and underlying mechanism. *Developmental Psychology, 35,* 303 – 317.

Schoner, G., & Thelen, E. (2006). Using dynamic field theory to rethink infant habituation. *Psychological Review, 113,* 273 – 299.

Scott, L. S. (2006). Featural and configural face processing in adults and infants. *Perception, 35,* 1107 – 1128.

Seeman, P., Bzowej, N. H., Guan, H. C., Bergeron, C., Becker, L. E., & Reynolds, G. P. (1987). Human brain dopamine receptors in children and aging adults. *Synapse, 1,* 399 – 404.

Sera, R., & Smith, L. B. (1990). New evidence on the development of the word "big." *Child Development, 61,* 1034 – 1052.

Seress, L., & Mrzljak, L. (1992). Postnatal development of mossy cells in the human dentate gyrus. *Hippocampus, 2,* 127 – 142.

Shields, P. J., & Rovee-Collier, C. (1992). Long-term memory for context-specific category information at six months. *Child Development, 63,* 245 – 259.

Shimizu, Y., & Johnson, S. C. (2004). Infants' attribution of a goal to a morphologically unfamiliar agent. *Developmental Science, 7,* 425 – 430.

Shtulman, A., & Carey, S. (2007). Improbable of impossible? How children reason about the possibility of extraordinary events. *Child Development, 78,* 1015 – 1032.

Shuwari, S. M., Albert, M. K., & Johnson, S. P. (2007). Discrimination of possible and impossible objects in infancy. *Psychological Science, 18,* 303 – 307.

Sitskoorn, M. M., & Smitsman, A. W. (1995). Infants' perception of dynamic-relations between objects: Passing through or support? *Developmental Psychology, 31,* 437 – 447.

Slater, A., Morison, V., & Somers, M. (1988). Orientation discrimination and cortical function in the human newborn. *Perception, 17,* 597 – 602.

Soken, N. H., & Pick, A. D. (1999). Infants' perception of dynamic affective expressions. *Child Development, 70,* 1275 – 1282.

Spelke, E. S. (1994). Initial knowledge. *Cognition, 50,* 431 – 445.

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science, 10,* 89 – 96.

Spelke, E. S., Vishton, P., & Von Hofsten, C. (1995). Object perception, object-directed action, and physical knowledge in infancy. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 165 – 180). Cambridge, MA: MIT Press.

Tan, L. S. C., & Bryant, P. (2000). The cues that infants use to distinguish discontinuous quantities: Evidence using a shift-rate recovery paradigm. *Child Development, 71,* 1162 – 1178.

Tang, Y., Zhan, W., Chen, K., Feng, S., Ji, Y., Shen, J., et al. (2006). Arithmetic processing in the brain shaped by cultures. *Proceedings of the National Academy of Sciences, 103,* 10775 – 10780.

Tanguay, P. E., Ornitz, E. M., Kaplan, A., & Bozzo, E. S. (1975). Evolution of sleep spindles in childhood. *Electroencephalography and Clinical Neurophysiology, 38,* 175 – 181.

Taylor, M. J., McCarthy, G., Saliba, E., & Degiovanni, E. (1999). ERP evidence of developmental changes in processing of faces. *Clinical Neurophysiology, 110,* 910 – 915.

Tinbergen, N. (1951). *The study of instinct.* London: Oxford University Press.

Uller, C., Carey, S., Huntley-Fenner, & Klatt, L. (1999). What representations might underlie infant numerical knowledge? *Cognitive Development, 14,* 1 – 36.

Valenstein, E. S. (2005). *The war of the soups and the sparks.* New York: Columbia University Press.

Walden, T., Kim, G., McCoy, C., & Karrass, J. (2007). Do you believe in magic? Infants' social looking during violations of expectations. *Developmental Science, 10,* 654 – 663.

Warrington, E. K., & McCarthy, R. A. (1987). Categories of knowledge. *Brain, 110,* 1273 – 1296.

Webb, S. J., Long, J. D., & Nelson, C. A. (2005). A longitudinal investigation of visual event-related potentials in the first year of life. *Developmental Science, 8,* 605 – 616.

Weikum, W. M., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastian-Galles, N., & Werker, J. F. (2007). Visual language discrimination in infancy. *Science, 316,* 1159.

Weiss, M. J., Zelazo, P. R., & Swain, I. U. (1988). Newborn response to auditory stimulus discrepancy. *Child Development, 59,* 530 – 541.

West, M. (1914). *Infant care. Bureau of publication 8.* Washington, DC: US Government Office.

Witter, M. P. (1993). Organization of the entorhinal-hippocampal system: A review of current anatomical data. *Hippocampus, 3,* 33 – 44.

Wynn, K. (1992). Addition and subtraction by human infants. *Nature, 358,* 749 – 750.

Wynn, K. (1998). Psychological foundations of number: Numerical competence in human infants. *Trends in Cognitive Science, 2,* 296 – 303.

Xu, F., Spelke, E. S., & Goddard, S. (2005). Number sense in human infants. *Developmental Science, 8,* 88 – 101.

Xu, Y., & Chun, M. M. (2007). Visual grouping in human parietal cortex. *Proceedings of the National Academy of Sciences, 104,* 18766 – 18771.

Child Development, November/December 2008, Volume 79, Number 6, Pages 1625–1632

Rediscovering Development in Infancy

Joseph J. Campos
University of California, Berkeley

David Witherington
University of New Mexico

David I. Anderson
San Francisco State University and University of California, Berkeley

Carl I. Frankel
Vanderbilt University and University of California, Berkeley

Ichiro Uchiyama
Doshisha University, Kyoto, Japan

Marianne Barbu-Roth
CNRS, Universite de Paris, IX

This commentary endorses J. Kagan's (2008) conclusion that many of the most dramatic findings on early perceptual, cognitive, and social competencies are ambiguous. It supports his call for converging research operations to disambiguate findings from single paradigms and single response indices. The commentary also argues that early competencies must be placed into a longitudinal framework, thereby allowing researchers to (a) identify whether regressive phenomena play a role in skill development, (b) understand what functions (if any) given skills play in their precocious manifestations and whether these functions are comparable in later instantiations of skills, and (c) avoid rich interpretations by identifying how robust a suspected competency is across contexts.

With eloquence and tact, Kagan (2008) has dropped a gauntlet, challenging the major intellectual and empirical trends in infancy research of the past quarter century. He has offered a persuasive rejoinder to those in developmental psychology who assign adult-like cognitive competencies to the young infant and who as a result both deemphasize and distract the field from the study of developmental transformations essential for understanding the baby. In so doing, Kagan goes well beyond correcting erroneous impressions about the adequacy of the evidence for early competencies in the infant. He argues heuristically, laying out implicitly and explicitly a number of major new lines of research using a fundamentally different approach to studying the human infant. These new lines of research involve supplementing the study of origins of psychological processes with descriptions of qualitative developmental transitions in infancy and adulthood and explanations of how those qualitative transitions take place.

Kagan has made his points by delineating the methodological and metatheoretical problems inherent in the work on infant perceptual/cognitive precocities. He identifies serious errors in the cornerstone assumption of work on precocity: the implicit belief that "impossible" events are more novel than "possible" ones and that the more novel an event is, the greater the looking time. Kagan reminds us that the functional relation of looking time to novelty is an inverted-U function; consequently, the duration of looking can be either *greater* to a novel stimulus than to a familiar, *equal* to one another, or *less*—depending upon where in the inverted-U function the novel and more familiar stimuli fall. This methodological critique renders ambiguous almost all studies linking "possible" and "impossible" events to lesser and greater looking times. To Kagan's critique, we add that the infant's registration that a transitional probability has been violated is an inadequate basis to infer that an infant has a theory of gravity or of object permanence, just as an infant's growing appreciation of differential phoneme transitional probabilities at word boundaries is an inadequate basis to infer that an infant has a theory of language. Much of the past quarter century of research in infancy needs to be

---

Preparation of this commentary was supported by National Science Foundation Grant BCS-0002001, National Institute of Child Health and Human Development Grants HD-39925 and HD-50638, Grant P20MD00262 from the National Center for Minority Health and Health Disparities, and a France-Berkeley Fund Award.

Correspondence concerning this article should be addressed to Joseph J. Campos, Institute of Human Development, University of California, Berkeley, CA 94708-1620. Electronic mail may be sent to jcampos@berkeley.edu.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7906-0004

reinterpreted, the issues about competencies need to be revisited, and widespread and incautious acceptance of claims for infant competencies need to be halted.

The ambiguity of inference about competencies is rendered yet more problematic by the paucity of converging research operations in "competent infant" research. Although the findings were never replicated, the methodology of Bower (1974), the father of the competent infant approach, provides exemplary illustration of converging research operations on the topic of early perceptual and cognitive competencies, especially depth perception. He claimed precocities in depth perception through the use of discriminated operant responding, avoidance reactions to looming stimuli, and reaching to a virtual object presented stereoscopically. Reporting findings that converged on consistent conclusions across paradigms, Bower provided a powerful case for precocious depth perception because these very different paradigms yielded the same conclusion. His methods stand as a textbook example of how an investigator should proceed to establish the validity of a conclusion about the presence and operation of an internal psychological process. Only the absence of replication kept his work from being definitive.

Sound converging research is necessarily more difficult than an initial research foray, so we are not surprised that competent infant research has made so few attempts at converging research and has not buttressed findings stemming largely from a single paradigm—habituation/dishabituation—and a single response index—one or another parameter of looking time. Consider, for example, the "violation of expectation" paradigm (e.g., Baillargeon, 1994; Wang, Baillargeon, & Brueckner, 2004). Despite alluding to the violation of expectancy paradigm as a "surprise" paradigm, this work never quantified the emotion of surprise facially, posturally, autonomically or with instrumental behaviors in the manner that Hiatt, Campos, and Emde (1979) did with older infants.

Studies of reaching in the dark (Clifton, Rochat, Litovsky, & Perris, 1991; Hood & Willatts, 1986; Robin, Berthier, & Clifton, 1996) show the advantage of converging research operations. These studies do yield some evidence for a rudimentary ability of 6-month-olds to keep in mind an absent object. However, even the work on reaching in the dark requires further converging research operations to hone in on object permanence, because. As conducted, such studies do not test the most important aspect of object permanence. Piaget's (1954) arguments on object permanence are intimately tied to means – ends behavior in relation to interference of the reach of a target by an obstacle or cover. In reaching in the dark, there is no intervening obstacle; there is no opportunity for the infant to place one action as subordinate (means) and a second one as superordinate (ends). Even if claims for precocious infant representational abilities ultimately prove true, it would be misleading on the evidence adduced to date to conclude that infants possess object permanence in the manner that is central to the theory that is supposedly being disproven. The same cautionary comment holds true for claims about rudimentary understanding of number concepts, imitation, or spatial awareness. At best, work to date on infant precocities has raised the possibility of competence in one skill (e.g., keeping in mind an absent entity for a few seconds) but has neither provided conclusive evidence for the existence of the precocity nor documented that the precocity plays a role in the more developed competency of the older child.

Kagan's critique goes well beyond documenting ambiguity of inference about competencies. The critique is also metatheoretical. He argues that the neglect of transformations evident in so much of the literature today incorrectly suggests a hypothetical continuity in development between early forms reported in the newborn or young child and later ones seen in the older toddler, child, and adolescent. He gives numerous instances of the difference between "schematic" levels of analysis that involve perception and simple memory, and semantic levels that imply greater organization and resistance to disruption by perceptual displays. His metatheoretical critique makes a powerful statement: Morphological similarities between early and later phenomena (such as "counting," imitation, or representation of hidden objects) do not imply similarity, let alone identity, of underlying process. Kagan insists that apparently similar earlier and later behaviors may mask different underlying mechanisms.

The present commentary builds on Kagan's article and points to new approaches to research. Our commentary depends heavily on the writings of Witherington (2007), Overton (1991), and Oppenheim (1981), whose publications on ontogenetic adaptations should be required reading for developmentalists. The main points we want to make are as follows:

1. There are continuities and discontinuities in development, losses as well as gains, reorganizations and transformations. The presentation of data from studies of a single age and a single paradigm is necessarily incomplete and, as we have noted before and will note again later, risk misleading the field.

2. The behavior of an infant at a given point/level of development cannot be assumed to provide evidence for a core competency leading to a well-developed final state. There may well be immediate functional consequences of behaviors shown at a particular phase of development that are adaptive *for that period*, without the competency having import for later behaviors. Additional evidence must be adduced that a competency gets built upon.
3. Development often consists of the hierarchical coordination of skills or component processes—the coordination yielding emergent outcomes. Such emergents represent reorganizations in form, wherein the emergent depends on prior competencies but does not "contain" them. The prior competencies take on qualitatively new, more general, and more flexible meanings.

Some Guidelines for Studying Development

In this section of our commentary, we interweave the three points just discussed into a set of guidelines designed to restore a developmental perspective to contemporary studies of infant competencies.

Guideline 1: Developmental Phenomena Are Embedded in Time: They Have a Forward Reference, a Contemporary Reference, and a Backward Reference

Development is not simply "change over time." To be truly developmental, the change must involve understanding how a phenomenon came about from the coordination of skills and processes present beforehand and how a phenomenon points to a future end state. To study a phenomenon developmentally, then, is to place it in the context of previous forms and potential subsequent forms (Oppenheim, 1981). Research on neonatal or precocious infant competencies is woefully lacking in this regard.

As we have already noted, contemporary competent infant researchers infer continuity in development from behaviors that are apparently morphologically similar at two very different points in time. This search for morphological similarities can lead to major errors. A good example of such an error is found in the work on neonatal imitation. The imitation seen in the newborn differs from that in older infants in many ways (see Jones, 2007, for an elaboration of this point), of which we discuss one: the restriction in scope (if not in existence; see Anisfeld et al., 2001) of behaviors like mouth opening and tonguing in neonates. Later, imitation is much more flexible, less stereotyped, and has a much broader "receptive field" than that of the newborn, extending well beyond simple oral behaviors. Imitation in the older infant occurs much more reliably and is much more robust to contextual variations, making later imitation far more capable of scaffolding communication. Moreover, to judge from Jones's (2007) recent report, neonatal imitative mouthing and tonguing largely disappear before the second half year of life, not to be seen again until 18 months of age and older. Where, then, is the continuity of a core competency for imitation and what happens to the parent–infant communication for which neonatal imitation is putatively the scaffold?

Is it preferable to conclude that there is indeed a loss of form and function between the weak and the ephemeral imitation that might exist in the newborn period and that shown in later infancy? Are the newborn skills still there, only remaining "hidden" (e.g., Thelen & Ulrich, 1991) or inhibited (Diamond, 1990)? Or, are there major transformations that take place in imitation across the first 2 years of life and thereafter evident only in as yet undone longitudinal studies designed to document what has been gained and what has been lost? We favor the third strategy, which is rich in potentially informative and unexpected findings, such as whether the development of imitation follows an ordinal sequence, and what mediates the broadening of the behaviors that an infant can imitate (see Uzgiris & Hunt, 1975).

Guideline 2: Context Affects the Breadth of Effect of Competencies

Research on infant competencies is notoriously weak in dealing with contextual factors that serve as boundary conditions for the manifestation of skills and create limitations of the generalizability of the behaviors. Consider the claim that 5-month-old infants possess the capacity to represent absent objects and to place two mental representations into spatial relation with each other (Baillargeon, 1994). This claim raises many more questions than it answers, questions that must be followed up to support the interpretation that the finding indicates a core competency. Otherwise, we have a mere snapshot in the developmental process—a snapshot that does not demonstrate how the infant at a given age arrived at the claimed precocity or how that phenomenon links up to later manifestations of "object permanence."

Let us review several of the outstanding issues that must be addressed before object permanence can be attributed to the 5-month-olds. Notice how constrained the testing paradigms used to demonstrate competencies are, how brief the demand is on the infant to "keep in mind" absent entities, and above all, as noted above, how lacking is description of how the precocious representation becomes linked to action later on, such as in means – ends behavior. Furthermore, the developmental background of the phenomenon has been lost; how is the object permanence of the 5-month-olds related to prior forms of reactions to objects that suddenly vanish? What is the mechanism by which precocious representation comes about? Do motoric skills play a role as inducers or facilitators of such representations, as they do at later ages (Kermoian & Campos, 1988)? For instance, does the new grasping skill of the 4- or 5-month-olds facilitate the ontogeny of representation because at this age, for the first time, a grasped object becomes occluded to sight but present to touch? We suspect that an analysis of the processes underlying early representations may reveal both how early representations are similar to and how they are different from, later and more complex representations. Such considerations about process are generally lacking. As a result, claims stemming from the study of precocious object permanence strike us as incomplete, even if we were to find the fundamental interpretation of looking time convincing.

Continuing with the theme of object permanence, we note how at older ages representation must become more enduring and more resistant to intrusions from new task demands. Impose an intermediate action between hiding and finding an object (i.e., make the representation serve a means – ends function), and the representation may become insufficient to guide behavior properly (Piaget, 1954). Change the posture of the infant between hiding and search and previously successful strategies may fall apart (Burger, 2004). The infant must have not only a representation of a hidden object but also the capacity to coordinate that representation with the act of removing a cover and then grasping the searched for object. Is it possible that the reported representational skill of the 5-month-olds is not sufficiently robust to mediate or play a role in these more cognitively demanding contexts? Is it possible that the memory of the 5-month-olds is weak and that any new task demand will disrupt the memory? And how does the 5-month-old's memory link up with the memory that mediates the complicated continuation of searching (lifting covers, looking all over the floor, shaking the container hiding the object, etc.) that a 14-month-old shows when he or she fails to find a toy where it has just been hidden (Ramsay & Campos, 1978)?

When viewed in this developmental light, the demonstration of the presence of a memory for spatial relations out of sight in a 5-month-old does seem to us strangely incomplete. The lacuna can be addressed in future research by adhering to the principle that processes typically have a precursor form and trend toward a more mature or well-developed end state. In addition, there must be recognition that contextual factors such as minimizing the task demands on the infant may facilitate the demonstration of apparent precocities and account for differences in performance by infants at different ages. Research keeping these points in mind is sorely needed. The 5-month-old's behavior, like that of the imitating newborn, must be placed into the developmental stream of time, with the investigator describing not only changes in behavioral outcomes with age but also (a) how behavior acquires a new level of organization with age or transformative events such as locomotion (Campos, 2006; Campos et al., 2000) and (b) how the infant's behavior trends toward a mature end state, described ahead of time. We also must understand how infants deal with perturbations they encounter in their environments and how alternative developmental pathways may be used to attain an end state (such as can occur in thalidomide infants; see Decarie, 1969; Kopp & Shaperman, 1973).

Guideline 3: There Are "Regressive" Phenomena in Development

No phenomena illustrate the need for embedding unsuspected competencies in developmental time better than those that show "regressive functions" in development. A regressive function refers to phenomena that are apparently present in the newborn and then become unobservable, and perhaps lost, only to emerge again at a higher level of function at a later age. These regressive phenomena include neonatal imitation, the newborn's grasp, the "empathic cry" of the newborn, postural compensation to patterns of optic flow, localization of sound, neonatal stepping, visual proprioception, and many others (Heimann, 2003). These regressive phenomena suggest that behaviors at one age that are apparently morphologically similar to those at later ages may in fact operate under rather different principles and may be controlled at one age by one neurophysiological process and at a later age by a different one (Gilmore, Hous, Petter, & Norcia, 2007). Alternatively, a set of behaviors can serve an adaptive function at one age, but the same or similar behaviors can serve a rather different function at a later age.

Regressive phenomena are well known to the developmentalist but are usually relegated to neonatal "reflexes." In fact, they apply to fairly mature reactions by the newborn. We will discuss two such regressive phenomena not usually considered as reflexes: the reach and grasping response and the postural response to peripheral optic flow.

Discontinuities in the reach and grasp response. When provided with sufficient head and trunk postural support, neonates engage in a form of reaching (actually "prereaching") behavior consisting of spontaneous forward arm extensions directed in the general direction of objects upon which visual attention is fixated. In the first few weeks of life, this prereaching activity is visually elicited, rather than visually guided. By that is meant that the newborn's reaction is triggered by vision (hence "elicited") but shows little behavioral flexibility (hence, not visually guided); by contrast, the behavior of the older infant shows modification of the path taken to the target of reach and an ability to adjust the response to different distal relations. The neonatal reach is typically accompanied by an opening of the hand, with fanlike extension of the fingers, but contact with objects is highly unreliable, with no attempt to grasp the objects (von Hofsten, 1982). Furthermore, hand opening in the context of prereaching acts as an extension of the arm movement synergy and not as an adaptive action in its own right; hand opening accompanies forward extension irrespective of whether the infant visually fixates an object. In agreement with the stance elaborated later in this commentary, von Hofsten (1982) has argued that the forward extension of reaching behavior in the newborn may differ from later developmental forms of reaching with respect to the functions it serves. Whereas reaching at 3–4 months is clearly in the service of *manipulation*, the earlier, visually elicited, reaches of the very young infant may function to *prepare the infant to orient or attend to, but not to manipulate*, an external object. Young infants' forward extensions, then, are functionally linked to looking as part of a general orienting response to objects. Thus, the forward extensions of 2-week-olds to an object may differ qualitatively in their functional significance relative to the same basic extensions (now integrated with grasping actions) of the 3-month-olds. In sum, that which appears morphologically similar at 1 week of age and 5 months of age proves to be functionally quite different across ages.

There is another important aspect of the reaching response that calls for caution among those who favor continuities in development. Over the course of the first 7 weeks of postnatal life, visually elicited reaches progressively decline in frequency as the young infant increases its visual inspection of objects, in the absence of any arm movements (von Hofsten, 1984). Between 2 and 3 months, however, reaching activity reemerges in frequency, but now forward extensions of the arm are accompanied by a fisted hand and characterized as swiping. After this period, around 4–5 months, infants begin again to open their hands as they reach for objects but now do so in the context of attempts to grasp objects that are visually fixated (von Hofsten, 1984). This visually guided or *goal-directed* reaching involves coordinated reaching and grasping activity and seems to be qualitatively distinct from the prereaching efforts that precede it. Neonatal reaching is thus not organized in the same way in the newborn and the 5-month-olds. Claiming continuities in this important skill of the child would be an error.

Discontinuities in visual proprioception. Another example of a phenomenon present in the neonate that becomes qualitatively transformed is the newborn infant's visual proprioception—the sense of self-motion produced by patterns of optic flow. Visual proprioception is what occurs when we experience induced and illusory self-motion when seated in a train: The movement of a train in the adjacent track creates a sense that one's self is moving and not the adjacent train. More generally, when there is optic flow moving toward a person in the visual periphery, the optic flow generates (a) a sense of self-motion in the direction opposite the movement of the optic flow and (b) a compensatory head, and later, total body, movement in the direction of the optic flow.

Newborns show backward head adjustments to optic flow moving toward the infant—an appropriate adaptive postural compensatory response to peripheral optic flow (Jouen, 1988). This demonstration is replicable (Jouen, 1990; Jouen, Lepecq, & Gapenne, 2000). However, as with reaching, the neonatal phenomenon appears to decline in effectiveness, only to return with renewed potency in the second half year of life. In the case of visual proprioception and peripheral optic flow, the catalyst for change proves to be locomotor experience.

Bertenthal and Bai (1989) documented that at 5 months of age, peripheral optic flow alone no longer elicits postural compensation, but it does at 9 months. Higgins, Campos, and Kermoian (1996) went on to show that crawling infants and prelocomotor infants of the same age having artificial "walker" experience respond appropriately to peripheral optic flow; prelocomotor infants of the same age as the other two groups do not—or do not do so as strongly. Furthermore, training in a powered mobility device brings about a heightened level of responsiveness to peripheral optic flow (Uchiyama et al., 2008). As with neonatal reaching, neonatal response to peripheral optic flow is evident, then may "hide" before the infant is 5 months of age, only to reappear at a higher level of organization at a later age in conjunction with the acquisition of another developmental skill—locomotion.

Guideline #4: Early Competencies May Serve Functions Quite Different From Later Ones

Oppenheim (1981) makes the point that a true developmental analysis requires the investigator to explore the contemporaneous functions that precocious organized behaviors serve. Our field honors in the breach this powerful developmental principle. We sometimes describe different levels of behavior in a skill domain, but we rarely speculate on, or test for, the adaptive function that each level of a skill has at one point in time, relative to the adaptive function that skill possesses later on.

One precocity in behavioral development, the so-called empathic cry of the newborn to the cry of another newborn, may illustrate Oppenheim's (1981) point well. As demonstrated in several labs and with different procedures, a newborn will cry to the wailing of another newborn but not to the cry of a chimpanzee, white noise, or the cry of an older child. As with neonatal reaching and visual proprioception, the empathic cry drops out, no longer evident at 5 months of age (Martin & Clark, 1982). Presumably, "empathic" reactions emerge again later, starting at 14 months of age, when infants show empathic instrumental behaviors to an adult "crying" in pain to a simulated injury (Dondi, Simion, & Caltran, 1999).

Our point in discussing the empathic cry is not simply to give one more illustration of regressive phenomena in development. Instead, our focus here is on the question: How does the neonatal cry to the cries of another infant have contemporary adaptive significance and how might the function of the cry for the newborn explain the presence of that skill early in life without there being a continuity with apparently similar processes occurring later on in life? Most contemporary studies on neonatal competencies do not consider what adaptive value the phenomenon of interest has at the age the infant is tested for the precocity. In the case of the neonatal cry, however, there is the possibility that a powerful adaptive function is being manifested by the infant crying to the cry of the other infant, but that the function is anything but empathic. Specifically, could it be that the function of the empathic cry is actually a competitive one? The possibility of such cries being competitive is premised on the fact that (a) crying by a neonate—any neonate—elicits the "let down" reflex of the mother, permitting nursing and (b) the cry of the target neonate can serve as a competitive call for the mother—now capable of nursing—to come and nurse her infant rather than someone else's infant. Proceeding with our argument, we wonder whether once the newborn period ends and a reciprocal relationship with the mother is established, the newborn's empathic cry loses its original competitive function and hence drops out of the infant's behavior repertoire. When an older child is later observed showing an empathic response to the mother or to another baby, this empathic behavior is a morphologically and functionally different reaction based on different developmental processes (Zahn-Waxler, Robinson, & Emde, 1992). In short, empathic phenomena in the neonate may have no relevance to empathy in the older child. Nevertheless, note that in our analysis, the empathic cry of the newborn does have major biologically adaptive significance in the neonatal period—as Oppenheim's (1981) principle would lead us to expect.

Concluding Comments

Clearly, this article is at best a beginning step toward a developmental analysis of behaviors that show apparent precocities. To talk about "concluding" is a bit like walking away from constructing a skyscraper when only a few floors have been erected. There is so much more to say about how one might proceed in taking a developmental perspective. For supplementation, we therefore recommend Witherington's (2007), Overton's (1991), and Oppenheim's (1981) more richly developed arguments. All three authors propose principles that are both more general and more abstract than those articulated here. As Witherington notes, developmental analysis begins with an understanding of the role of formal and final cause, as complements to the ready acceptance by developmentalists of efficient and material causes. When we explain behavior in terms of a particular stage or developmental level of organization, when we talk about the "organization of behavior" (e.g., Sroufe, 1996), we are appealing to formal cause—namely, how different skills are patterned in the here and now. When we talk about end points of development, we are appealing to final causes. Developmentalists do not take formal and final causes very seriously in the current state of the field, and as a result, we believe that our field suffers. To quote Murray (1991), "The earlier steps of stages in a developmental sequence of behaviors are made more intelligible by our knowledge of the end-point toward which they are progressing, even though the later periods cannot in any way be the efficient cause of the earlier events because of the unidirectional character of time" (p. 241). To

Murray's quote, we would add that we need to understand how behaviors are organized at each point in development, how they become progressively better coordinated, and how they function in the here-and-now adaptation of the individual (formal cause). A developmental perspective in its broadest sense then must consider both the increasingly complex patternings of behavior and the end states toward which development trends. It is instructive that current epigenetic approaches attempt just such a synthesis of material and efficient causes nested within final and formal ones. Epigenetic approaches call on all four levels of causality to fully explain how a phenomenon comes about with age.

We presented guidelines to help structure research toward a more organic and less static portrayal of development. In closing, we reiterate in different words the broad implications of our guidelines. These are that (a) many newborn and apparently precocious competencies may have nothing to do functionally with the mature competency and indeed may have no *contemporary* adaptive function at all (just as the lungs of the fetus are prepared for, but not typically called to, function), (b) some newborn competencies may be only a part of what will eventually be the mature one, and (c) a newborn competency may be thoroughly plastic (like a stem cell) and may be functional for different reasons at different points in the developmental stream. For all three reasons, assumptions about "core competencies" cry out for longitudinal investigations, which have been sorely lacking, and analyses of their contemporaneous and subsequent functions, which often are subject to neglect. As Kagan (2008) noted, the past quarter century of research on infant developmental precocities must be revisited and likely revised.

References

Anisfeld, M., Turkewitz, G., Rose, S., Rosenberg, F., Faigi, R., Sheiber, F., et al. (2001). No compelling evidence that newborns imitate oral gestures. *Infancy, 2*, 111–122.

Baillargeon, R. (1994). How do infants reason about the physical world? *Current Directions in Psychological Science, 3*, 133–140.

Bertenthal, B. I., & Bai, D. L. (1989). Infants' sensitivity to optical flow for controlling posture. *Developmental Psychology, 25*, 936–945.

Bower, T. (1974). *Development in infancy.*. San Francisco: W.H. Freeman.

Burger, S. (2004). Demands on finite cognitive capacity cause infants' perseverative errors. *Infancy, 5*, 217–238.

Campos, J. (2006, June). *Epigenetic processes in infant development.* Presidential address presented at the biennial meetings of the International Conference on Infant Studies, Kyoto, Japan.

Campos, J., Anderson, D., Barbu-Roth, M., Hubbard, E., Witherington, D., & Hertenstein, M. (2000). Travel broadens the mind. *Infancy, 1*, 149–219.

Clifton, R., Rochat, P., Litovsky, R., & Perris, E. (1991). Object representation guides infants' reaching in the dark. *Journal of Experimental Psychology: Human Perception and Performance, 17*, 323–329.

Decarie, T. G. (1969). A study of the mental and emotional development of the thalidomide child. In B. M. Foss (Ed.), *Determinants of infant behaviour IV* (pp. 167–188). London: Methuen.

Diamond, A. (1990). The development and neural bases of memory functioning as indexed by the AB and delayed response tasks in human infants and infant monkeys. *Annals of the New York Academy of Sciences, 608*, 263–317.

Dondi, M., Simion, F., & Caltran, G. (1999). Can newborns discriminate between their own cry and the cry of another newborn infant? *Developmental Psychology, 35*, 418–426.

Gilmore, R., Hous, C., Petter, M., & Norcia, A. (2007). Development of cortical responses to optic flow. *Visual Neuroscience, 24*, 845–856.

Heimann, M. (2003). *Regression periods in human infancy.* Mahwah, NJ: Erlbaum.

Hiatt, S., Campos, J., & Emde, R. (1979). Emotional expression in infancy: A study of facial response patterns. *Child Development, 50*, 1020–1035.

Higgins, C., Campos, J., & Kermoian, R. (1996). Effect of self-produced locomotion on infant postural compensation to optic flow. *Developmental Psychology, 32*, 836–841.

Hood, B., & Willatts, P. (1986). Reaching in the dark to an object's remembered position: Evidence for object permanence in 5-month-old infants. *British Journal of Developmental Psychology, 4*, 57–65.

Jones, S. (2007). Imitation in infancy: The development of mimicry. *Psychological Science, 18*, 593–599.

Jouen, F. (1988). Visual-proprioceptive control of posture in newborn infants. In B. Amblard, A. Berthoz, & F. Clarac (Eds.), *Posture and gait: Development, adaptation and modulation* (pp. 59–65). Amsterdam, Netherlands: Elsevier Science.

Jouen, F. (1990). Early visual-vestibular interactions and postural development. In H. Bloch & B. Bertenthal (Eds.), *Sensory-motor organizations and development in infancy and early childhood* (pp. 199–215). Dordrecht, Netherlands: Martinus Nijhoff.

Jouen, F., Lepecq, J., & Gapenne, O. (2000). Optic flow sensitivity in neonates. *Infant Behavior and Development, 24*, 271–284.

Kagan, J. (2008). In defense of qualitative changes in development. *Child Development, 79*, 1606–1624.

Kermoian, R., & Campos, J. (1988). Locomotor experience: A facilitator of spatial cognitive development. *Child Development, 59*, 908–917.

Kopp, C., & Shaperman, J. (1973). Cognitive development in the absence of object manipulation during infancy. *Developmental Psychology, 9*, 430.

Martin, G., & Clark, R. (1982). Distress crying in neonates: Species and peer specificity. *Developmental Psychology, 18*, 3–9.

Murray, F. B. (1991). Questions a satisfying developmental theory would answer: The scope of a complete explanation of development phenomena. In P. van Geert & L. P. Mos (Eds.), *Annals of theoretical psychology* (Vol. 7, pp. 237–247). New York: Plenum Press.

Oppenheim, R. (1981). Ontogenetic adaptations and retrogressive processes in the development of the nervous system and behaviour: A neuroembryological perspective. In K. J. Connolly & H. F. R. Prechtl (Eds.), *Maturation and development: Biological and psychological perspectives* (pp. 73–109). London: Heinemann.

Overton, W. F. (1991). The structure of developmental theory. In P. van Geert & L. P. Mos (Eds.), *Annals of theoretical psychology* (Vol. 7, pp. 191–235). New York: Plenum Press.

Piaget, J. (1954). *The construction of reality in the child*. New York: Basic Books.

Ramsay, D., & Campos, J. (1978). The onset of representation and the infant's entry into stage VI of object permanence development. *Developmental Psychology, 14*, 79–86.

Robin, D., Berthier, N., & Clifton, R. (1996). Infants' predictive reaching for moving objects in the dark. *Developmental Psychology, 32*, 824–835.

Sroufe, L. A. (1996). *Emotional development: An organizational perspective*. New York: Cambridge University Press.

Thelen, E., & Ulrich, B. (1991). Hidden skills: A dynamic systems analysis of treadmill stepping during the first year of life. *Monographs of the Society for Research in Child Development, 52*(1, Serial No. 223).

Uchiyama, I., Anderson, D., Campos, J., Witherington, D., Barbu-Roth, M., Frankel, C., et al. (2008). Locomotor experience affects self and emotion. *Developmental Psychology, 44*, 1225–1231.

Uzgiris, I., & Hunt, J. (1975). *Ordinal scales of infant development*. Urbana: University of Illinois Press.

von Hofsten, C. (1982). Eye-hand coordination in the newborn. *Developmental Psychology, 18*, 450–461.

von Hofsten, C. (1984). Developmental changes in the organization of prereaching movements. *Developmental Psychology, 20*, 378–388.

Wang, S., Baillargeon, R., & Brueckner, L. (2004). Young infants' reasoning about hidden objects: Evidence from violation-of-expectation tasks with test trials only. *Cognition, 93*, 167–198.

Witherington, D. C. (2007). The dynamic systems approach as metatheory for developmental psychology. *Human Development, 50*, 127–153.

Zahn-Waxler, C., Robinson, J., & Emde, R. (1992). The development of empathy in twins. *Developmental Psychology, 28*, 1039–1047.

Child Development, November/December 2008, Volume 79, Number 6, Pages 1633 – 1638

In Defense of Core Competencies, Quantitative Change, and Continuity

Paul C. Quinn
University of Delaware

J. Kagan (2008) urges contemporary developmentalists to (a) be cautious when attributing conceptual knowledge to infants based on looking-time performance, (b) constrain their interpretation of infant performance with multiple methodologies, and (c) reconsider the possibility that qualitative development may be the path by which perceptual infants become conceptual adults. This commentary outlines an account of conceptual development that adheres to two of the three Kagan provisos. It is (a) circumspect in the core competencies attributed to infants and (b) grounded in convergent measures including looking time, event-related potentials, computational modeling, and eye tracking, but (c) maintains that the transition from the perceptually based category representations of infants to the knowledge-rich concepts of adult is a continuous developmental process marked by quantitative change.

In a provocative article, Kagan (2008) has rebooted the vigorous debate that sprang up about a decade ago when Haith's (1998) article "Who Put the Cog in Infant Cognition? Is Rich Interpretation Too Costly?" was published by *Infant Behavior & Development*. The debate centers on the issue of what kinds of skills and knowledge can be attributed to infants based on their performance in experiments conducted with looking-time methodologies. Kagan contends that one is unlikely to observe adult competencies in infants and argues that looking-time results offer multiple interpretations, which need to be verified with convergent methodologies. Kagan also argues that the extent to which infant performance falls short of adultlike levels of competence should point investigators toward reembracing the notion of qualitative change as the path by which the nature of infant thought may be altered to attain the conceptual power of adult knowledge.

I agree with the argument that whatever mental structures are guiding infant performance in looking-time experiments, their content is unlikely to match the richness of the mature semantic concepts underlying adult competence. In addition, I agree with the argument that the multiple interpretations of infant looking time need to be evaluated further with convergent methodologies. It also seems reasonable to suggest that a vocabulary for describing infant performance be adopted that respects the immaturity of the infant's mental structures relative to those of adults. However, to develop this vocabulary in such a way that it emphasizes dissociation between infants and adults may not be warranted and to contend that qualitative change is the path to the end state of being fully developed may not be necessary. The remainder of the commentary will propose an account of early conceptual development that avoids rich interpretation of infant looking-time performance and uses multiple methodologies to further constrain that interpretation. But the account also argues that much of conceptual development can be accounted for by core competencies that when put into operation in the developing infant can begin to yield functional knowledge in a short period of time and that a continuous process of developmental enrichment marked by quantitative change may explain the transition from infant to child to adult concepts.

Perceptually Based Category Representations in Infants

Looking-time studies measuring the emergence of categorization during early development have demonstrated that young infants between the ages of 3 and 7 months will categorize animals such as cats, dogs, and horses (e.g., Mareschal & Quinn, 2001). The infants are presented with multiple instances from a common category during familiarization (e.g., cats) and then with a preference test pairing a novel instance from the familiar category (e.g., a new cat) with a novel instance from a novel category (e.g.,

---

Preparation of this commentary was supported by National Institutes of Health Grants HD-42451 and HD-46526.

Correspondence concerning this article should be addressed to Paul C. Quinn, Department of Psychology, University of Delaware, Newark, DE 19716. Electronic mail may be sent to pquinn@udel.edu.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7906-0005

a new dog). Categorization is inferred if infants generalize their familiarization to the novel instance from the familiar category and display a preference for the novel instance from the novel category.

Categorization Versus Discrimination

It is worth emphasizing that the animal stimuli presented to the infants were realistic photographic images. In the case of cats, the stimuli depicted different breeds in a variety of colors and stances. In fact, the stimuli from within a category were highly discriminable, an observation that underscores that the novel category preferences cannot be attributed merely to processes of perceptual discrimination between categories. The infants can discriminate between the different instances of the cats just as readily as they can discriminate between a cat and a dog, yet they are also able to look through this variation so as to group together the various cat images into a common representation that excludes the dog images. As Kagan (2008) notes, "Young infants and many animal species can discriminate between pictures of an elm tree and an elephant and between parallel and nonparallel lines. But this evidence does not mean that they possess the semantic concepts of plant, animal, or parallel" (p. 1609). I agree completely with Kagan's observation (is there anyone who would disagree?) but would also note that the categorization task presents infants with two novel stimuli in the preference test, thereby calling on infants to generalize from the familiarized instances to a novel instance from the same category and distinguishing the categorization task from a simple discrimination task.

Perceptually Based Category Representations Versus Perceptual Schemata

Being careful not to overinterpret infant performance, I have adopted the stance Kagan recommends in describing the mental structure guiding infant looking behavior with a construct that differs from the term that might be used to describe adult performance. That is, infant performance has been described as being mediated by "perceptually based category representations" rather than "concepts." This is to respect the fact that the infants at issue are just a few months of age, and they are responding to visual images presented in the laboratory. They do not "know" about cats and dogs in the way that adults know about their more abstract genetic and reproductive attributes (e.g., cats have cat DNA and give birth to kittens). At the same time, I see no reason to minimize the accomplishments of the infants by relegating the status of their representations to "perceptual schemata," thereby creating a dissociation between the terminology used to describe the performance of infants and adults participating in category learning tasks. As described by Kagan, "Schemata are used primarily to recognize events, places, and objects experienced in the past" (p. 1609). By contrast, an important feature of category representations, even if based on perceptual information, is that they allow the infant to "go beyond the information given" in terms of being able to generalize to novel instances not previously experienced.

A Convergent Measure

As a complement to the behavioral looking-time studies demonstrating category formation for the various animal species by infants, Quinn, Westerlund, and Nelson (2006) asked whether neural correlates of category learning by infants might be measurable. They did so by developing an event-related potential (ERP) analogue of the behavioral looking-time procedure. The infants were presented with 36 cat images followed by 20 novel cat images interspersed with 20 novel dog images. The results (i.e., the neural activations observed in different parts of the scalp up to a second and a half after stimulus presentation) were partitioned into four different conditions: Cats 1–18, Cats 19–36, novel cats, and novel dogs. The rationale is that if there is a neural signal that corresponds with category learning, then the responses to Cats 1–18 and the novel dogs (reflecting initial experience with exemplars of a category) should be equivalent. The responses to Cats 19–36 and the novel cats (reflecting a learned category of cats) should also be equivalent but different from the responses to Cats 1–18 and the novel dogs.

Of interest was a late slow wave component observed in the time window between 1,000 and 1,500 ms after stimulus presentation. In studies of recognition memory for individual stimuli in infants (Nelson, 1994; Reynolds & Richards, 2005), the amplitude of the slow wave has been associated with the differentiation of familiar and novel stimuli. In particular, a slow wave that returns to baseline has been associated with recognition of familiarity, whereas a slow wave that deflects away from baseline in the negative direction is associated with detection of novelty. The results from Quinn et al. (2006) were that a negative slow wave was observed over left occipital parietal scalp in the time window between 1,000 and 1,500 ms after stimulus onset in response to Cats 1–18 and novel dogs, whereas a slow wave that returned to baseline was observed for Cats 19–36 and novel cats. This analysis reveals that the infant's brain responded to novel cats with activity equivalent to that displayed for Cats 19–36. More generally, it points to the neural instantiation of a key behavioral indicant of categorization: responding to the novel as if it is familiar. That there is distinct brain activity corresponding to the formation of a category representation for the exemplars presented during familiarization (e.g., the cats) provides convergent evidence with the looking-time studies and indicates a neural preparedness in the first half year of life to represent category information on the basis of perceptual experience.

Quantitative Versus Qualitative Change

The studies of infant learning of animal categories suggest that young infants divide the world of objects into perceptual clusters that later come to have conceptual significance for adults. As such, the conceptual representations found later in development may be viewed as informational enrichments of the category representations that infants form on the basis of perceptual experience (Quinn & Eimas, 1997). For example, infants who are presented with exemplars of cats and horses are not experiencing these exemplars as an undifferentiated mass but rather as separate groups that fall into distinct representations. These representations might then serve as placeholders for the acquisition of the more abstract information that occurs beyond infancy, through language and more formal learning of the nonobvious attributes associated with semantic categories (Quinn & Eimas, 2000). Thus, over time, the perceptual placeholder representation for cats will come to include the information that cats eat tuna, hunt mice, have cat DNA, give birth to kittens, and are labeled as "cats," whereas the representation for horses will come to include the information that horses eat hay, carry heavy loads, have horse DNA, give birth to foals, and are labeled as "horses." The acquisition of this additional information serves to enrich the original perceptually based category representations to the point that they attain the richness of the more mature conceptual representations of children and adults (Quinn, 2004). By this view, what changes as concepts mature is the content of the representations rather than the processes underlying their development (Madole & Oakes, 1999; Rakison & Poulin-Dubois, 2001).

We have also been careful not to overinterpret our category learning studies as indicating that infants are forming concepts that they leave the lab with and then carry around in their heads for everyday usage. The claim is that infants are demonstrating perceptual grouping or clustering abilities in the laboratory that are presumably engaged when infants encounter cats and horses in the course of experience in the world outside the laboratory (or images of them as depicted in videos and picture books). The latter representations are those believed to be the ones that serve as the supports for the acquisition of the nonobvious attributes that occurs beyond the infancy period.

Against Dissociating the Perceptual From the Semantic

I agree with Kagan's (2008) major claim that "perceptual competences" can be distinguished from "conceptual structures," but I also disagree with the assertion that qualitative change need be the path to get from one to the other. More generally, I disagree with the idea that the "physical features" of events should be separated from their "meaning." Even the most conservative theorists of cognitive development would grant that infants have operational perceptual input systems and a general learning mechanism to acquire a database of information about objects in the world. And that knowledge may be perceptual in nature. In the case of animals, stored data may include information about faces, coloring, skeletal appendages, a body shape bounded by curved contours, movement patterns, and species-specific sounds of communication. Kagan is strangely silent on the idea that powerful learning mechanisms provided with such perceptual input can acquire knowledge previously thought to be part of the infant's conceptual endowment (Elman et al., 1996). Such mechanisms, instantiated in the form of connectionist learning systems that provide yet another source of convergent evidence for those interested in explicating the mental status of infants, have provided important in-principle demonstrations that a single network (and hence a single system of representation) can produce a broad-to-narrow trajectory for category learning that was once thought to provide a distinctive signature for learning concepts separately from percepts (Quinn, 2002; Quinn & Johnson, 1997; cf. Mandler, 2000). Notably, the ERP evidence described earlier in which infants were simply provided with perceptual experience with visual images of cats and dogs also produced neural evidence for learning the more global category (cat + dog) prior to separating cats and dogs into distinct groupings (Quinn et al., 2006).

These observations raise the question of whether perceptual information should be dissociated from semantic categorization as Kagan and others have contended (Carey, 2000; Mandler, 2000). An alternative view (one that I embrace) is that knowledge about the perceptible parts and properties of objects can be semantic knowledge (see also Murphy, 2002, in press). One could not have much of a concept of cats, for example, without knowing what they looked like and what parts they had. Concepts must include perceptual information or else they would not be very helpful. It is hard to imagine how a child could even map the more abstract attributes onto their correct object referents without having category representations available from perceptual experience to serve as support structures. Even schoolchildren learning about biology must be able to recognize the Birman lying on the sofa, the long-haired gray Maine Coon sitting in the window across the street, and the short-haired orange tabby that just began patrolling in the yard next door as instances of "cats."

By this view, it becomes difficult to envision that the concepts formed by infants from perceptual experience are simply cast aside when a later concept that is imbued with more abstract information becomes available. It is important to acknowledge that this view does not deny that in some instances, more abstract information can supplant perceptual information when the two sources of information are placed in competition as determinants of category identity (Keil, 1989; Plunkett, Hu, & Cohen, 2008). This type of evidence has been used to support an "essentialist" view of concept development (Gelman, 2003). However, although "essentialist" arguments have been forceful in the literature, Hampton, Estes, and Simmons (2007) have actually presented cases in which a majority of research participants weight characteristic perceptual features more heavily than defining essentialist features in decisions about category identity, thereby suggesting that at least in some instances information about perceptual properties is considered just as semantic as information about genetics. Moreover, if one accepts the argument that perceptual information may be meaningful, then the difference between perceptually and conceptually based concepts is less clearly a qualitative one. Indeed, in nature, perceptual (form) and conceptual (function) information are often correlated and experiments designed to tease apart whether perceptual or conceptual information is the more potent determinant of categorization behavior may serve only to perpetuate what is arguably a false dichotomy. Instead, the way forward may be to explain how perceptual and conceptual knowledge are integrated (not dissociated) to form mature concepts (Murphy, 2002, in press; Quinn, 2004).

A Role for Biases

The account of category development offered in the present commentary, like that offered in Kagan's (2008) target article, has attempted to avoid radical nativist solutions that rely on innate knowledge, but it does recognize that there are likely to be constraints or preexisting biases of a general nature that help make the rapid development of categorization possible. Biases may assist infants in solving the nontrivial problem of determining which features should be selected for determining category membership. And while the evidence reviewed indicates that young infants can learn category representations on the basis of perceptual experience, it does not inform us as to the particular features the infants use to form the category representations. The answer to this question is not obvious, given that the exemplars from the various nonhuman animal categories possess a number of common attributes such as a head, torso, four legs, and a tail.

Quinn and Eimas (1996) examined infants' abilities to categorize cats versus dogs when provided with the whole stimuli, just the heads (with the bodies occluded) or just the bodies (with the heads occluded) during both the familiarization and the preference test portions of the experiment. The results were that the infants categorized when presented with either the whole stimuli or just the heads but not when presented with just the bodies. These findings indicate that the head provided a sufficient basis for the infants to categorize cats versus dogs. The conclusion of Quinn and Eimas (1996) was bolstered by an additional study in which infants were familiarized with whole cat or dog images and then preference tested with hybrid stimuli (i.e., cat head on dog body vs. dog head on cat body; Spencer, Quinn, Johnson, & Karmiloff-Smith, 1997). Infant preference during test followed the direction of the novel category head. Interestingly, adults also seem to emphasize the head when representing animal species (Reed, McGoldrick, Shackelford, & Fidopiastis, 2004).

Use of the head could arise from a biasing mechanism that directs infant attention to the head information that is present in a visual scene (Johnson & Morton, 1991). It is also possible that the infants simply learn during the course of experience with the exemplars that the head is the most diagnostic part of the stimulus. In other words, the head feature may be flexibly created as the basis for the category representation in an online fashion (Quinn, Schyns, & Goldstone, 2006). To determine whether infants used the head because of a bias or because of online learning, Quinn, Doran, Reiss, and Hoffman (in press) used an eye-tracking methodology to measure the eye fixations of infants as they scanned the visual images presented in the categorization task. On the assumption that stimulus regions used for categorization will be preferentially fixated over those not used, it was reasoned that if the head preference results from a biasing mechanism, then infants should fixate more on the heads than on the bodies of the exemplars throughout the course of the familiarization portion of a category learning procedure. Alternatively, if the infants learn that the head is the most diagnostic region of the stimuli, then the head preference should emerge during the course of the familiarization trials. The major finding was that throughout the course of the familiarization trials, the infants were shown to have a marked preference for fixating the head that was present during the whole of familiarization.

The results are consistent with the idea that the reliance on the head for the categorical parsing of cats and dogs arises from a biasing mechanism. However, one could still argue that infants are simply orienting to the head because it contains high contrast internal features, and infants are attracted to the most visible portion of the stimulus (Banks & Salapatek, 1981). To address this possibility, a control study was undertaken that repeated the eye-tracking version of the category familiarization procedure but with stimulus images that were inverted (Quinn et al., in press). If the head preference resulted from infants' simply orienting to the most visible portion of the stimulus, then it should still be observed with inversion. However, if the infants orient to the head because of a bias that is in place to facilitate face recognition, then one would not expect it to be present with stimulus inversion, given that inversion changes the normal configuration of facial features, and faces are believed to be recognized on a configural basis. The major finding from the control study was that fixations to the head region of the stimuli were small when compared with fixations to the body.

The inverted control results provide evidence that infants use the head to categorize upright cat and dog images because of a biasing mechanism that responds to face information. This type of bias could facilitate conceptual development by allowing infants to differentiate categories that have faces (e.g., animals) and those that do not (e.g., furniture), as well as partition classes marked by distinctive facial make ups (e.g., cats vs. dogs). More generally, a bias of this nature may aid infants in selecting from among various features that are potentially available in the input (i.e., head or body), and in this way "set the system on the trajectory of learning" (Thelen & Smith, 1994, p. 315). Such biases may be especially important in determining the course of concept acquisition in a system that is otherwise characterized by flexibility.

Concluding Comments

This commentary has outlined an account of concept development that adheres to the guidelines set forth by Kagan (2008) in terms of being (a) cautious about the skills and knowledge attributed to infants and (b) based on evidence that converges from multiple methodologies (i.e., looking time, ERPs, computational modeling, and eye tracking). However, in contrast with Kagan, the account maintains that the perceptually based category representations of infants should not be dissociated from the knowledge-rich concepts of adults. Rather, the latter grow out of the former based on a continual process of enrichment in which category representations that are initially based on perceptual attributes come to incorporate nonobvious attributes acquired through informal and formal tuition and language. From this perspective, infants may begin to form categories based on core competencies consisting of functioning perceptual systems (including that for language) and a general learning mechanism (that can represent within-category similarity and between-category dissimilarity) that is facilitated in its operation by biases to attend to some inputs more so than others. Such competencies equip infants with the ability to form functional category representations that, although not adultlike in their initial form, will become adultlike during continuous real-time development that occurs in day-to-day interactions with a structured environment. Infants are endowed with the tools to build a foundation for conceptual development, rather than with a preconceptual form that will be shed once metamorphosis provides mature concepts.

References

Banks, M. S., & Salapatek, P. (1981). Infant pattern vision: A new approach based on the contrast sensitivity function. *Journal of Experimental Child Psychology, 31*, 1–45.

Carey, S. A. (2000). The origin of concepts. *Journal of Cognition and Development*, *1*, 37–42.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press.

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. Oxford, UK: Oxford University Press.

Haith, M. M. (1998). Who put the cog in infant cognition? Is rich interpretation too costly? *Infant Behavior & Development*, *21*, 167–179.

Hampton, J. A., Estes, Z., & Simmons, S. (2007). Metamorphosis: Essence, appearance, and behavior in the categorization of natural kinds. *Memory & Cognition*, *35*, 1785–1800.

Johnson, M. H., & Morton, J. (1991). *Biology and cognitive development: The case of face recognition*. Oxford, UK: Blackwell.

Kagan, J. (2008). In defense of qualitative changes in development. *Child Development*, *79*, 1606–1624.

Keil, F. C. (1989). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.

Madole, K. L., & Oakes, L. M. (1999). Making sense of infant categorization: Stable processes and changing representations. *Developmental Review*, *19*, 263–296.

Mandler, J. M. (2000). Perceptual and conceptual processes in infancy. *Journal of Cognition and Development*, *1*, 3–36.

Mareschal, D., & Quinn, P. C. (2001). Categorization in infancy. *Trends in Cognitive Sciences*, *5*, 443–450.

Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.

Murphy, G. L. (in press). What are categories and concepts? In D. Mareschal, P. C. Quinn, & S. E. G. Lea (Eds.), *The making of human concepts*. Oxford, UK: Oxford University Press.

Nelson, C. A. (1994). Neural correlates of recognition memory in the first postnatal year of life. In G. Dawson & K. Fischer (Eds.), *Human behavior and the developing brain* (pp. 269–313). New York: Guilford Press.

Plunkett, K., Hu, J. F., & Cohen, L. B. (2008). Labels can override perceptual categories in early infancy. *Cognition*, *106*, 665–681.

Quinn, P. C. (2002). Early categorization: A new synthesis. In U. Goswami (Ed.), *Blackwell handbook of childhood cognitive development* (pp. 84–101). Oxford, UK: Blackwell.

Quinn, P. C. (2004). Multiple sources of information and their integration, not dissociation, as an organizing framework for understanding infant concept formation. *Developmental Science*, *7*, 511–513.

Quinn, P. C., Doran, M. M., Reiss, J. E., & Hoffman, J. E. (in press). Time course of visual attention in infant categorization of cats versus dogs: Evidence for a head bias as revealed through eye tracking. *Child Development*.

Quinn, P. C., & Eimas, P. D. (1996). Perceptual cues that permit categorical differentiation of animal species by infants. *Journal of Experimental Child Psychology*, *63*, 189–211.

Quinn, P. C., & Eimas, P. D. (1997). A reexamination of the perceptual-to-conceptual shift in mental representations. *Review of General Psychology*, *1*, 271–287.

Quinn, P. C., & Eimas, P. D. (2000). The emergence of category representations during infancy: Are separate perceptual and conceptual processes required? *Journal of Cognition and Development*, *1*, 55–61.

Quinn, P. C., & Johnson, M. H. (1997). The emergence of perceptual category representations in young infants: A connectionist analysis. *Journal of Experimental Child Psychology*, *66*, 236–263.

Quinn, P. C., Schyns, P. G., & Goldstone, R. L. (2006). The interplay between perceptual organization and categorization in the representation of complex visual patterns by young infants. *Journal of Experimental Child Psychology*, *95*, 117–127.

Quinn, P. C., Westerlund, A., & Nelson, C. A. (2006). Neural markers of categorization in 6-month-old infants. *Psychological Science*, *17*, 59–66.

Rakison, D. H., & Poulin-Dubois, D. (2001). Developmental origin of the animate-inanimate distinction. *Psychological Bulletin*, *127*, 209–228.

Reed, C. L., McGoldrick, J. E., Shackelford, R., & Fidopiastis, C. (2004). Are human bodies represented differently from other animate and inanimate objects? *Visual Cognition*, *11*, 523–550.

Reynolds, G. D., & Richards, J. E. (2005). Familiarization, attention, and recognition memory in infancy: An ERP and cortical source localization study. *Developmental Psychology*, *41*, 598–615.

Spencer, J., Quinn, P. C., Johnson, M. H., & Karmiloff-Smith, A. (1997). Heads you win, tails you lose: Evidence for young infants categorizing mammals by head and facial attributes (Special Issue: Perceptual Development). *Early Development and Parenting*, *6*, 113–126.

Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge, MA: MIT Press.

Child Development, November/December 2008, Volume 79, Number 6, Pages 1639 – 1647

Defending Qualitative Change: The View From Dynamical Systems Theory

John P. Spencer
University of Iowa and Iowa Center for Developmental and Learning Sciences

Sammy Perone
University of Iowa

A central controversy in developmental science, enflamed by nativist accounts, is whether development is best viewed as a series of qualitative or continuous changes. This article defends the notion of qualitative change from the perspective of dynamical systems theory (DST). Qualitative change within DST refers to the shift that occurs when a system goes from one attractor state through an instability into a different attractor state. Such changes occur on the second-to-second timescale of behavior. Thus, developmental analysis must always stay local, grounded in the real-time attractor states around which behavior is organized. This article also demonstrates that qualitative and continuous change should not be cast in opposition. They are aligned concepts that work together across multiple timescales.

In his article entitled "In Defense of Qualitative Changes in Development," Kagan (2008) argues that the balance between claims of transformational change and continuity in development has been a central point of controversy throughout the long history of our science. This controversy has reached a head in the past two decades given the growing number of claims of continuity between infants' abilities and the abilities of older children. Many of these claims, Kagan argues, stem from a nativist tradition that has built its reputation, in part, by demonstrating amazing infant abilities and their seeming connection to later forms of cognition (e.g., Meltzoff & Moore, 1977; Wynn, 1992). Kagan argues that such claims of continuity are exaggerated. For instance, the ability of children to use complex semantic networks is qualitatively different from the perceptual schemata formed by infants in looking tasks. Kagan also argues that the empirical evidence of early competence is on tenuous ground given that the empirical record is largely built from a single measure of infant performance—total looking time—which is known to be subject to a host of influences that are often overlooked by researchers interested in showing early competence.

We applaud Kagan's effort to defend the notion of qualitative change. Our article seeks to do the same; however, we take a different tack motivated by our commitment to a particular theory of development—dynamical systems theory (DST). DST was introduced to developmental science in the early 1990s with the publication of *A Dynamic Systems Approach to the Development of Cognition and Action* by Thelen and Smith (1994; for related ideas, see van der Maas & Molenaar, 1992; Van Geert, 1998). These authors, inspired by earlier work on DST in physics, motor control, and psychology, proposed a radical view of development as fluid, multidetermined, and grounded in perception and action (for reviews, see Spencer et al., 2006; Thelen & Smith, 2006). In the past two decades, the concepts of DST have been applied to a host of phenomena including topics in motor development (e.g., Corbetta & Thelen, 1996), cognitive development (e.g., Spencer, Simmering, Schutte, & Schöner, 2007), and socioemotional development (e.g., Lewis, Lamey, & Douglas, 1999).

In the present article, we use this theoretical framework to weigh in on the issue of whether development is best viewed through a qualitative or continuous lens. Our article has two central goals. First, we will define what "qualitative" means from a dynamical systems perspective. Critically, DST provides a precise definition of this concept. This is important because the notion of qualitative development means different things to different researchers, creating conceptual confusion in the literature. Second, we will argue that qualitative changes in behavioral organization actually occur at the timescale of behavior—the second-to-second timescale of thoughts, actions, and emotions. From this view, what changes over learning and development is the emergence of these qualitatively different real-time behavioral states, their stability, and the ability of the infant or child to flexibly shift from one state to another.

We would like to thank Gregor Schöner and Larissa Samuelson for helpful discussions and comments on an earlier version of this article. This work was supported by National Science Foundation HSD 0527698 awarded to J.P.S.

Correspondence concerning this article should be addressed to John P. Spencer, 11 Seashore Hall E, University of Iowa, Iowa City, IA 52242. Electronic mail may be sent to john-spencer@uiowa.edu.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7906-0006

What Does Qualitative Mean?

For many developmental scientists, the term *qualitative* invokes the notion of transformational change—creating something new that was not obviously connected to the something old. A typical example is the transformation that occurs when a caterpillar emerges from a cocoon as a butterfly. What went into the cocoon bears only faint resemblance to what came out. This seems to capture the use of qualitative in the article by Kagan. He highlights that the perceptual schemata hypothesized to underlie infants' behavior bear no obvious resemblance to the semantic networks invoked in childhood.

Although the use of the term *qualitative* in this context has appeal due to its face validity, there is an important tension point—we know that there must be some fundamental, biological connection between the caterpillar and the butterfly. Thus, must there not be continuity as well as discontinuity or transformation? Must the perceptual schemata not be somehow connected to the emergence of the first semantic network?

To further compound the problem, there is now a rich body of evidence showing that apparently abrupt and transformational changes previously thought to reflect qualitative reorganizations in development appear more continuous when examined at finer levels of detail (Adolph, Robinson, Young, & Gill-Alvarez, 2008; Siegler, 1994, 1996). This work has highlighted that the measurement scale and level of analysis can critically effect the conclusions we draw from data about the nature of developmental change, as well as our ability to test theories that make claims about transformational and continuous developmental processes.

Given the diversity of thought regarding the meaning of the term *qualitative,* how this concept relates to developmental continuity and how to measure different types of change, we think the time is ripe to anchor the notion of qualitative development to a detailed theoretical framework. DST has much to offer in this regard: DST offers a precise definition of qualitative change, clear empirical signatures to identify qualitative change, and clarity regarding the link between qualitative and continuous change. Although DST has much to offer, it is difficult to explain dynamical systems concepts succinctly and in a way that resonates with the interests of behavioral researchers. In an effort to achieve both goals, we describe a set of central concepts below using examples that link neural and behavioral dynamics in real time. We then extend these concepts to tackle the issue of qualitative *developmental* change.

Dynamical systems are a class of systems that can be mathematically modeled using equations that specify the rate of change of some variable or system of variables at each moment in time. The idea behind this class of equations is actually quite powerful: If you know the current state of the system and its rate (and direction) of change, that is, how fast the system is moving along some trajectory, then you can take the current state, add the amount you expect it to change, and predict where the system will be at the next point in time.

If we could do this perfectly, we could describe—in detail—the development of a child and predict the state of the child each step along the way. The problem is that the dynamical systems we study in development are extraordinarily complex. Because we cannot know all the details of a complex system, we have to settle for something simpler—we have to study how the system behaves around special points called attractors, points toward which the system gravitates over time. This can get tricky, though, because these special points can actually change themselves. This is what occurs when a qualitative transition happens in a dynamical system: The system shifts from being in the local territory of one attractor to losing that attractor and gravitating toward a new one. The question is: How do we know when one attractor becomes unstable and another attractor emerges (which is formally called a bifurcation; see Braun, 1994)?

Figure 1 provides an overview that answers this question. This figure also illustrates how difficult it can be to identify qualitative shifts in the dynamics of a system. Figures 1A to 1C show a set of stimulus events—inputs to a hypothetical brain of a hypothetical child—turning on and off from second to second as might occur in an infant habituation experiment (see, e.g., Schöner & Thelen, 2006). These figures also show the activation of one simple neuron in the child's brain. Note that in all the simulation examples in this article, we will talk about an example neuron to pick the simplest case possible that is still directly relevant to behavior. Importantly, however, all the concepts we introduce are applicable to the more complex neural fields described in many of our articles (e.g., Simmering, Schutte, & Spencer, 2008; Spencer & Schöner, 2003; Spencer et al., 2007). Moreover, these dynamics can be found in real cortical fields (Amari, 1977; Amari & Arbib, 1977) and measured in real brains using electrophysiology (Bastian, Riehle, Erlhagen, & Schöner, 1998; Bastian, Schöner, & Riehle, 2003; Erlhagen, Bastian, Jancke, Riehle, & Schöner, 1999; Jancke et al., 1999).

As can be seen in Figure 1A, the first exemplary neuron shows a rather complex pattern of behavior:

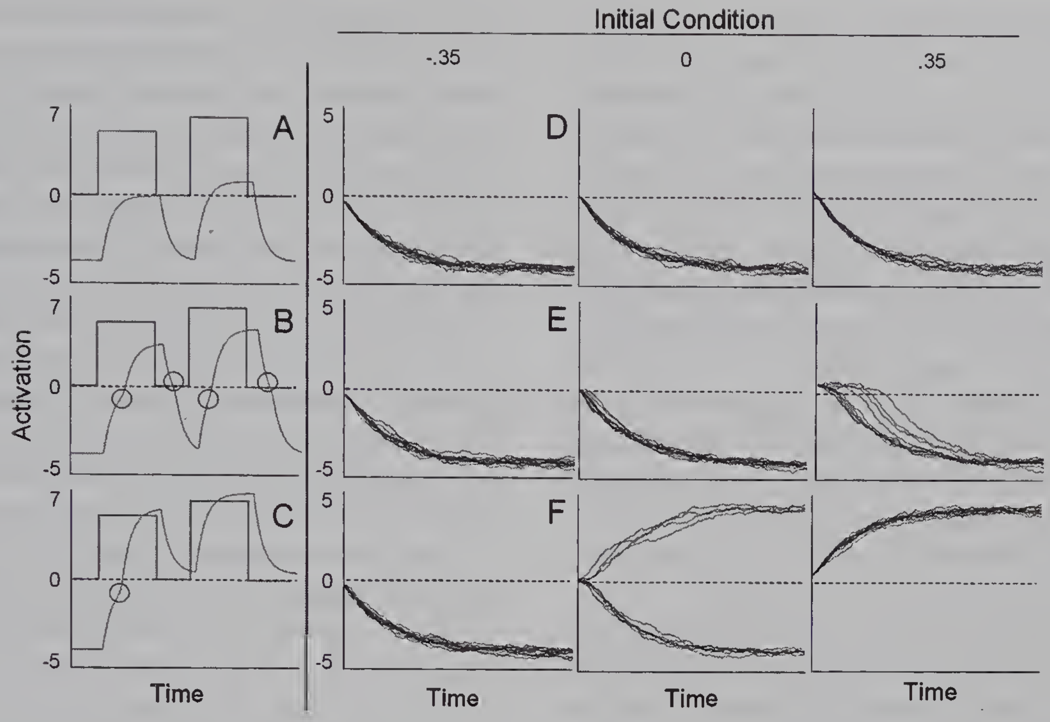


Figure 1. Simulations of a simple neural system to illustrate the concept of qualitative change within dynamical systems theory: (A) the behavior of a neuron with linear dynamics in response to input, (B) the behavior of a self-stabilized neuron, and (C) the behavior of a self-sustaining neuron.

Note. Rectangular inputs are shown in gray. Circles denote qualitative transitions in neural activation. Panels in (D), (E), and (F) show repeated simulations of the neurons simulated in (A), (B), and (C) with different initial activation values (−0.35, 0, or 0.35) and a small amount of noise.

Its activation rises and falls with an exponential pattern as each stimulus is presented, and the overall level of activation covaries with the stimulus strength. Given the nonlinear behavior of this neuron, one might guess that a nonlinear dynamical system is at work that goes up toward one attractor state when the input is on and down toward a qualitatively different state when the input is off. And because the details of the attractor state shift as the input shifts, one might invoke the idea that that system rises to a more advanced or richer encoding of the stimulus when the second input is presented.

It turns out that this analysis is only partially correct. The neuron in Figure 1A is moving toward a higher level of activation when the input is on and a lower level of activation when the input is off. Critically, however, there is no qualitative shift in the dynamics of this system whatsoever. Rather, the activation profile shown in Figure 1A was produced by a simple *linear* dynamical system with a *single* attractor state that the neural activation is moving toward across the duration of this simulation. Why does the neural activation go up and down? This occurs because the location of the attractor in activation space moves as the input is turned on and off.

In contrast to Figure 1A, the simulation shown in Figure 1B shows an example of a qualitative shift. In particular, this neuron shifts from a *resting* or off attractor state into an input-dominated attractor state we refer to as the *self-stabilized* state. In this attractor state, activation stays stably on (i.e., above 0) provided that the stimulus is present, and the strength of neural activation within this state can accurately mirror the strength of the input signal (see Figure 1B). If the input is removed for a significant period of time, however, the self-stabilized state becomes unstable and the system moves through another qualitative shift as it returns to the resting state. Indeed, there are four qualitative changes in the dynamics of this neural activation profile across the simulation shown in Figure 1B (see circled joints in the activation profile).

Conceptually, the self-stabilized state reflects a type of perceptual encoding: The shift into the self-stabilized state marks a detection decision, that is, the neuron has detected the presence of a reliable signal and it will remain in this state provided that the input remains reasonably robust. For instance, this neuron can stay in the self-stabilized state even when there are transient interruptions in the input signal as might occur when an infant moves her eyes briefly to look away from a stimulus and back again (for discussion, see Spencer, Perone, & Johnson, in press).

It turns out that the simple neuron simulated in Figure 1B can actually move into a third attractor state—the self-sustaining state—shown in Figure 1C.

In this simulation, the neural activation moves from the resting state into a state where activation stays stably on (above 0) even when the input is removed. Thus, there is only one qualitative shift in the behavior of the neuron in Figure 1C (see circled joint). Once the neuron detects the stimulus, it maintains an active representation of this detection decision for the duration of the simulation. The neuron is able to enter this attractor state because it has a relatively strong self-excitation parameter. Consequently, once stimulated, the neuron is able to keep excitation around, even when the external stimulus is turned off. This strong self-excitation is evident in the strength of activation in Figure 1C: At two points in this simulation, the strength of neural activation actually rises to a level that exceeds the input strength.

Just as the self-stabilized state provides a conceptual link to notions of perceptual encoding, the self-sustaining state captures a neurally plausible form of working memory (Fuster, 1994; Simmering et al., 2008; Spencer et al., 2007; Wang, 2001). We have shown, for example, that this basic form of sustained neural activity underlies the ability of children and adults to actively maintain information in spatial working memory during short-term delays (Spencer & Hund, 2003; Spencer et al., 2007). More recently, we have shown that this form of working memory also underlies the maintenance of nonspatial features such as color and orientation (Johnson, Spencer, Luck, & Schöner, 2008; Johnson, Spencer, & Schöner, 2008). And developmental changes in the ability to stably enter this working memory state can capture changes in the stability of spatial working memory over development (Schutte, Spencer, & Schöner, 2003; Simmering et al., 2008), the likelihood that infants and young children will perseverate in simple reaching tasks (Schutte et al., 2003; Spencer, Smith, & Thelen, 2001; Thelen, Schöner, Scheier, & Smith, 2001), as well as developmental changes in the capacity of the visual working memory system (Simmering & Spencer, 2008).

What determines whether a neuron or neural field is stabilized by input (Figure 1B) versus self-sustaining (Figure 1C)? The move between these two states requires a tiny tweak to the neuron's self-excitatory strength. With weaker self-excitation, the neuron enters the stabilized state; with stronger excitation, the neuron can be self-sustaining. This highlights that there is a deep connection between the notions of continuous change and transformational change within DST: A continuous and subtle change in the details of a neural system can produce a qualitative transformation in the attractor states the system moves into and out of in real time. Indeed, this tie between continuity and transformation becomes more intricate when one considers that the neural attractor states shown in Figures 1B and 1C are emergent and softly assembled from several factors—even in this simplest of cases. One can, for instance, move the neuron into the self-sustaining state (rather than the self-stabilized state) by boosting the strength of the input (for discussion, see Simmering et al., 2008).

Thus far, we have largely just asserted that the attractor states illustrated in Figures 1A to 1C are qualitatively different. One signature of the qualitative shifts are the joints in the neural activation profiles in Figures 1B and 1C that occur near an activation value of 0 as the system moves from the off state into a stable on state. In Figures 1D to 1F, we go one step further highlighting another key signature of behavior as a dynamical system moves through a qualitative transition—heightened variability near the transition point (van der Maas & Molenaar, 1992; Van Geert, 1998). In these simulations, we set the starting state of the neuron's activation just below 0, at 0, and just above 0, and then let neural activation change through time in the absence of input, but in the presence of a small bit of neural noise. What emerges are rather dramatic differences in how the linear system (Figure 1D), the self-stabilized neuron (Figure 1E), and the self-sustaining neuron (Figure 1F) behave.

As can be seen in Figure 1D, the linear system shows the same behavior regardless of its starting state—in the absence of input, neural activation shows exponential decay to a baseline level. This highlights that there is no special qualitative transition happening in this system as activation passes through the zero threshold. By contrast, Figure 1E behaves in a qualitatively different fashion once the initial activation level is greater than 0 (see right panel in Figure 1E). Now, self-excitation is strong enough that the system can hover in an on state temporarily, even in the absence of stimulus input. Note that if we were to compute the variability of neural activation across the simulations in this panel, the variance would be dramatically higher than in all the simulations shown in Figure 1D. The final set of simulations shown in Figure 1F shows the most dramatic case of heightened variability near a qualitative transition: When we set the neuron's initial activation state to 0 (see middle panel of Figure 1F), the system shows a roughly 50–50 chance of going into the off versus the on state.

Considered together, the simulations in Figure 1 illustrate what the concept of qualitative change means within DST. Three points are critical. First, the notion of qualitative change in DST is all about a change in the attractor states of a system—a shift from one stable state, through an instability, into a different attractor state. Second, you cannot judge a qualitative transition by its cover, that is, by just watching the system behave through time. Even simple systems with no qualitative changes in dynamics (a single attractor state) can exhibit complex patterns of behavior through time. To discover whether a qualitative transition has occurred, one must look deeper for signatures of transformational change such as high variance near transition points. Third, there is a deep connection between continuity and qualitative change within DST: qualitative transformations can arise from continuous changes in multiple aspects of the system, such as when a small change in self-excitation moves a neural system from a perceptual to a working memory state.

What About Qualitative Transitions in Development?

Although the dynamical behavior of the simple neuron in Figure 1 is interesting, it is a far cry from the types of developmental transitions typically studied in developmental science. What does this notion of qualitative change from DST buy us when thinking about qualitative *developmental* change?

Developmental changes in the dynamics of a system are often depicted as a shifting landscape of attractor wells. Figure 2 shows an early depiction of this notion from Waddington (1954), as well as an updated view from Muchisky, Gershkoff-Stowe, Cole, and Thelen (1996). In these figures, the organism is represented by the ball moving along a landscape. As each new valley opens up over development, qualitatively new states are available to the organism. For instance, the wells in Figure 2B capture different categories of action that emerge as infants shift from reaching, to crawling, to walking, and so on. Thus, the collection of wells at some moment in development captures the landscape of potential actions. And the depth of each well captures the stability of that particular action type (a deeper well is more stable).

Although this depiction of development is pervasive, there is an odd thing about depicting development in this manner. As deeper wells form as a function of skill acquisition, practice, consolidation, and so on, the system—the ball—will have a harder time getting out of the most stable states. This is a good thing if we are talking about the stability of an adult's walking pattern on bumpy ground, but it is not a good thing if we are talking about an adult's ability to flexibly shift from one stable action pattern (walking) to another (dancing). Skill involves stability (deep wells), but it also involves the ability to destabilize one pattern and move into another. Skill, therefore, involves qualitative change. How does the ball manage that nifty trick?

It turns out that the developmental landscape picture is a bit off in this regard. Although one can think of the ball getting kicked from one well to another to capture movement from one pattern to another, this does not effectively capture skilled behavior. Why? Empirical evidence shows that flexible, skilled behavior involves moving from one stable attractor state through an instability into another stable attractor state (see, e.g., Scholz & Kelso, 1989). This is precisely what our neuron did in Figure 1. Although one can think of this as having two wells and moving a ball back and forth, the more accurate picture is that the ball is in one well, *then that well*

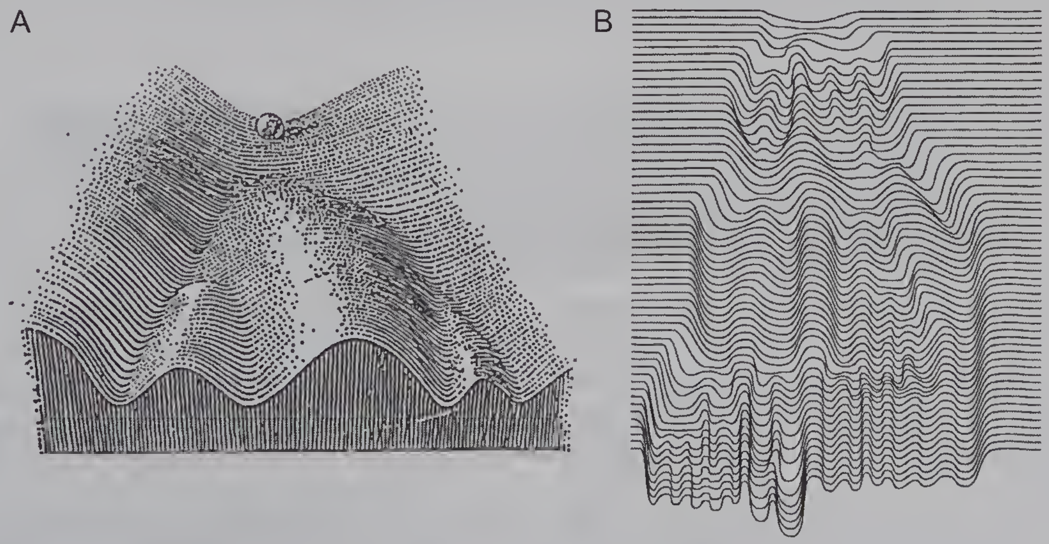


Figure 2. Depictions of developmental landscapes that invoke the concepts of dynamical systems theory: (A) epigenetic landscape described by Waddington (1954) and (B) developmental landscape described in Muchisky, Gershkoff-Stowe, Cole, and Thelen (1996).

disappears, and the ball rolls to a newly formed well nearby. By this view, attractor states are not sitting around; they are not competencies waiting to be tapped. Rather, attractors come and go in real time—the attractor landscape is dynamic and changing on the timescale of behavior (for related ideas, see Spivey, 2007).

It is important to note that the developmental landscapes in Figure 2 are not wrong; they are just limited. They provide a temporally averaged view of the real dynamics that are taking place. This can offer a useful survey of the terrain, but it can blur some important details, in this case, that the qualitative shifts that are taking place in development actually come and go and emerge and morph in real time.

How can we put the notion of real-time qualitative shifts (see Figure 1) together with a developmental landscape (see Figure 2)? Figure 3 shows one view of this integration, returning to our simple neural example from the previous section. The first thing we have done in this simulation is run it for many, many time steps—41,000 here compared to 400 in Figure 1. Thus, this figure is showing real-time changes in the behavior of our neuron (or neural field) over a timescale that might encompass a day or two of experience.

What is the neuron experiencing? Figure 3A shows the input to the neuron. The duration of the inputs vary to reflect some of the natural variations an infant might encounter as she moves from context to context. We have also simulated a simple form of developmental change. In particular, Figure 3B shows a slow, continuous (linear) change in the strength of self-excitation. This type of change might arise from a simple Hebbian process that strengthens synaptic connections that are repeatedly active (for related ideas, see Mareschal et al., 2007; Schutte et al., 2003; Simmering et al., 2008; Spencer, Dineva, & Schöner, in press).

As can be seen in Figure 3C, the combination of variations in input, random fluctuations in activation, and changes in self-excitation lead to the emergence of qualitatively new attractor states over the course of development in this simulation. At the start of the simulation, the neuron lives in a resting state (see the solid line above the neural activation profile). It shows some fluctuations up and down that mimic the input pattern, but it never enters the self-stabilized state.

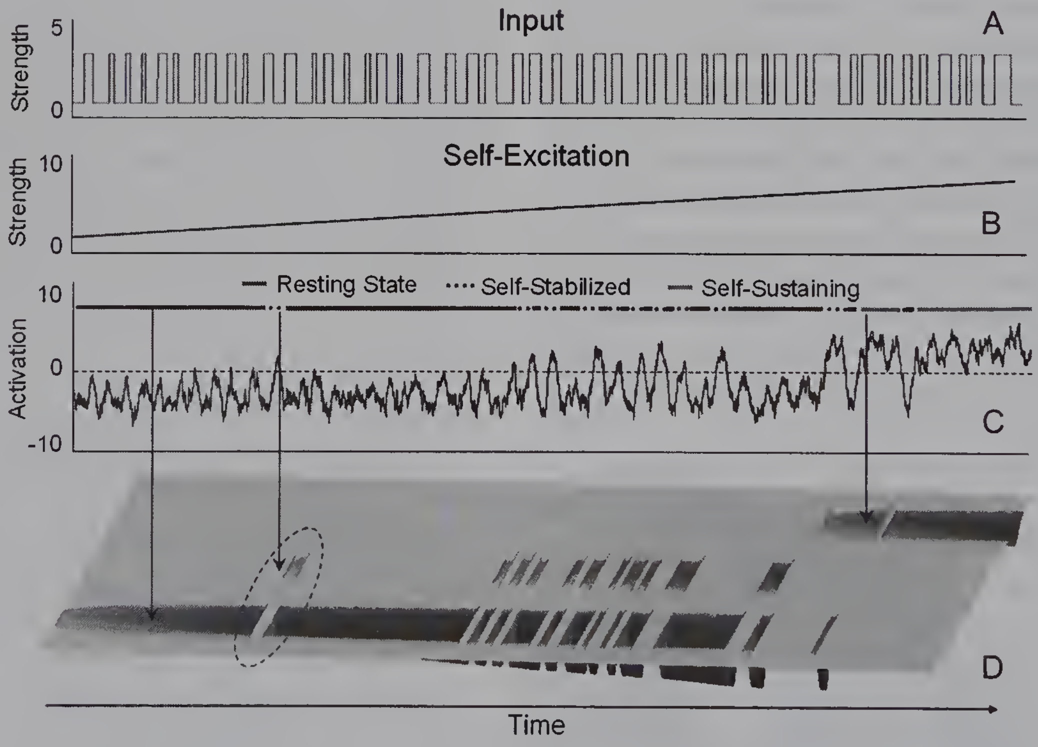


Figure 3. Simulation of a simple neural system in real time across a long time span to illustrate the integration of timescales within dynamical systems theory: (A) pattern of rectangular inputs over time, (B) linear increase in the neuron's self-excitation strength over time, (C) activation of the neuron over time, and (D) a developmental landscape constructed by mapping the attractor states shown above the activation profile in (C) onto a well diagram.
Note. Thick line (C) above the neural activation profile indicates the attractor state the neuron is in at each moment in time (solid line = resting state; dotted line = self-stabilized state; gray line = self-sustaining state). Arrows (D) show example of mapping for each attractor state. Circled transition marks the first qualitative transition in real time from the resting to the self-stabilized state. Depth of each well is scaled based on cumulative frequency of each attractor state over time.

As development unfolds, however, there is a qualitative shift in the neuron's behavior: It becomes more and more likely that the neuron will enter the self-stabilized state where it encodes the presence of a stimulus (see dotted line above the activation profile). Critically, although the neuron spends more time in the self-stabilized state, it is still moving into and out of this state through real time. Toward the end of this long simulation, the system first enters the self-sustaining or working memory state (see gray line above the activation profile). Again, we emphasize that this state first emerges in real time in the context of a particular stimulus input; however, over time, the system more reliably enters this state as self-excitation strengthens.

Figure 3D shows a developmental landscape that we created by transforming the categorization of the neuron's attractor states through time (see solid, dotted, and gray lines in Figure 3C) into a well diagram. Note that this landscape looks different from the images in Figure 2, in part, because time runs from left to right rather than from top to bottom. But there is also a more critical difference in how the wells change over time. Rather than having the emergence and deepening of new wells over time (Figure 2), *a qualitative change in the attractor layout occurs each time the neuron goes from one state to another in real time* (the first such occurrence is marked by the circle in Figure 3D). Over development, new states emerge in real time (first the self-stabilized state, then the self-sustaining state), these states become more stable (i.e., deeper), and the states are entered more reliably.

In summary, Figure 3 shows how qualitative shifts in behavior emerge in real time and cohere over the longer timescales of learning and development to give rise to the attractor states commonly depicted in developmental landscapes. Critically, Figure 3 clarifies how multiple timescales can be integrated within a developing system, making static wells dynamic in real time.

Implications of DST for Developmental Science

What are the implications of a dynamical systems view of qualitative change for developmental science? First, there is something very special about the notion of qualitative change within DST. This concept is about creating something new—a new attractor state—out of something old—the existing, diverse, and complex components that are newly assembled. It is about the emergence of new forms in real time as infants and children behave in real situations and how these new forms come to repeatedly cohere over learning and development.

Second, the integration of dynamics over multiple timescales highlights that developmental stability and change is always connected to the local details of behavior. Consequently, studying qualitative shifts in behavior over broad stretches of time—while often useful as a starting point—can sometimes reveal relatively little about the organization of behavior in real time. In this context, we agree with Kagan that claims of continuity across broad stretches of time that attempt to link infant behaviors with the behaviors of much older children must be treated as placeholders—approximations that should always be viewed with a critical eye. Importantly, however, this same critique applies whether we extend the arrow of time forward in development from infancy to childhood or backward in development from infancy to the prenatal period and purported biological (or evolutionary) predispositions (Spelke, 1998; Spelke & Kinzler, 2007). Approximations that extend far in both temporal directions should be treated with much more care in our science (Spencer, Blumberg, et al., in press).

The third implication of a dynamical systems view is that understanding qualitative developmental shifts is extremely complicated. One cannot just observe the local behavior of a system and extract the attractor states around which behavior is organized. Rather, one must look for the unique signatures that separate one state from another. And given that attractor states are softly assembled and multiply determined, it can be difficult to understand why they come and go, especially when the components that make up behavior are linked in nonobvious ways. These behavioral realities—which have been born out in many well-studied empirical examples—provide even more reason to think about development locally and be highly dubious of claims of long-term continuity, nativist or otherwise (see Spencer, Blumberg, et al., in press).

In conclusion, the challenge of a dynamical systems view is that to understand development, we must understand how the dynamics of a system change over multiple timescales, how qualitatively new attractor states emerge in real time, and how those attractors become more accessible, more stable, and more tuned to the particulars of the situation over development. Thus, we agree with a central theme of Kagan's (2008) article: Qualitative shifts in development occur, and they should be defended because they are fundamental to what emergence over development is all about. But qualitative change should not be put in opposition to gradual, continuous change as Kagan's article implies. Continuity and change go hand in hand. We think DST offers a firm defense of qualitative development, a formal framework for understanding the link between qualitative development and continuity, and a charge to push our understanding of development to the more local timescale where emergence and continuity come together to truly create something new out of something old.

References

Adolph, K. E., Robinson, S. R., Young, J. W., & Gill-Alvarez, F. (2008). What is the shape of developmental change? *Psychological Review, 115*, 527–543.

Amari, S. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics, 27*, 77–87.

Amari, S., & Arbib, M. A. (1977). Competition and cooperation in neural nets. In J. Metzler (Ed.), *Systems neuroscience* (pp. 119–165). New York: Academic Press.

Bastian, A., Riehle, A., Erlhagen, W., & Schöner, G. (1998). Prior information preshapes the population representation of movement direction in motor cortex. *Neuroreport, 9*, 315–319.

Bastian, A., Schöner, G., & Riehle, A. (2003). Preshaping and continuous evolution of motor cortical representations during movement preparation. *European Journal of Neuroscience, 18*, 2047–2058.

Braun, M. (1994). *Differential equations and their applications* (4th ed.). New York: Springer Verlag.

Corbetta, D., & Thelen, E. (1996). The developmental origins of two-handed coordination: A dynamic perspective. *Journal of Experimental Psychology: Human Perception & Performance, 22*, 502–522.

Erlhagen, W., Bastian, A., Jancke, D., Riehle, A., & Schöner, G. (1999). The distribution of neuronal population activation (DPA) as a tool to study interaction and integration in cortical representations. *Journal of Neuroscience Methods, 94*, 53–66.

Fuster, J. M. (1994). *Memory in the cerebral cortex*. Cambridge, MA: MIT Press.

Jancke, D., Erlhagen, W., Dinse, H. R., Akhavan, A. C., Giese, M., Steinhage, A., et al. (1999). Parametric population representation of retinal location: Neuronal interaction dynamics in cat primary visual cortex. *Journal of Neuroscience, 19*, 9016–9028.

Johnson, J. S., Spencer, J. P., Luck, S. J., & Schöner, G. (2008). *A dynamic neural field model of visual working memory and change detection*. Manuscript submitted for publication.

Johnson, J. S., Spencer, J. P., & Schöner, G. (2008). Moving to a higher ground: The dynamic field theory and the dynamics of visual cognition. *New Ideas in Psychology, 26*, 227–251.

Kagan, J. (2008). In defense of qualitative changes in development. *Child Development, 79*, 1606–1624

Lewis, M. D., Lamey, A. V., & Douglas, L. (1999). A new dynamic systems method for the analysis of early socioemotional development. *Developmental Science, 2*, 457–475.

Mareschal, D., Johnson, M. H., Sirois, S., Spratling, M., Thomas, M., & Westermann, G. (2007). *Neuroconstructivism: Vol. I. How the brain constructs cognition*. Oxford, UK: Oxford University Press.

Meltzoff, A. N., & Moore, M. K. (1977). Imitation of facial and manual gestures by the human neonate. *Science, 198*, 75–78.

Muchisky, M., Gershkoff-Stowe, L., Cole, E., & Thelen, E. (1996). The epigenetic landscape revisited: A dynamic interpretation. In C. Rovee-Collier & L. P. Lipsitt (Eds.), *Advances in infancy research* (Vol. 10, pp. 121–159). Westport, CT: Ablex.

Scholz, J., & Kelso, J. A. S. (1989). A quantitative approach to understanding the formation and change of coordinated movement patterns. *Journal of Motor Behavior, 21*, 122–144.

Schöner, G., & Thelen, E. (2006). Using dynamic field theory to rethink infant habituation. *Psychological Review, 113*, 273–299.

Schutte, A. R., Spencer, J. P., & Schöner, G. (2003). Testing the dynamic field theory: Working memory for locations becomes more spatially precise over development. *Child Development, 74*, 1393–1417.

Siegler, R. S. (1994). Cognitive variability: A key to understanding cognitive development. *Current Directions in Psychological Science, 3*, 1–5.

Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking*. New York: Oxford University Press.

Simmering, V. R., Schutte, A. R., & Spencer, J. P. (2008). Generalizing the dynamic field theory of spatial cognition across real and developmental time scales. *Brain Research, 1202*, 68–86.

Simmering, V. R., & Spencer, J. P. (2008). *Developing a magical number 7 minus 4: The dynamic field theory reveals why visual working memory capacity estimates differ across tasks*. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th annual conference of the Cognitive Science Society* (pp. 1537–1542). Washington, DC: Cognitive Science Society.

Spelke, E. S. (1998). Nativism, empiricism, and the origins of knowledge. *Infant Behavior & Development, 21*, 181–200.

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science, 10*, 89–96.

Spencer, J. P., Blumberg, M. S., McMurray, B., Robinson, S. R., Samuelson, L. K., & Tomblin, J. B. (in press). Short arms and talking eggs: Why we should no longer abide the nativist-empiricist debate. *Child Development Perspectives*.

Spencer, J. P., Clearfield, M., Corbetta, D., Ulrich, B., Buchanan, P., & Schöner, G. (2006). Moving toward a grand theory of development: In memory of Esther Thelen. *Child Development, 77*, 1521–1538.

Spencer, J. P., Dineva, E., & Schöner, G. (in press). Moving toward a grand theory while valuing the importance of the initial conditions. In J. P. Spencer, M. S. Thomas, &

J. L. McClelland (Eds.), *Toward a unified theory of development: Connectionism and dynamic systems theory re-considered*. New York: Oxford University Press.

Spencer, J. P., & Hund, A. M. (2003). Developmental continuity in the processes that underlie spatial recall. *Cognitive Psychology, 47*, 432–480.

Spencer, J. P., Perone, S., & Johnson, J. S. (in press). The dynamic field theory and embodied cognitive dynamics. In J. P. Spencer, M. Thomas, & J. L. McClelland (Eds.), *Toward a unified theory of development: Connectionism and dynamic systems theory re-considered*. New York: Oxford University Press.

Spencer, J. P., & Schöner, G. (2003). Bridging the representational gap in the dynamical systems approach to development. *Developmental Science, 6*, 392–412.

Spencer, J. P., Simmering, V. R., Schutte, A. R., & Schöner, G. (2007). What does theoretical neuroscience have to offer the study of behavioral development? Insights from a dynamic field theory of spatial cognition. In J. M. Plumert & J. P. Spencer (Eds.), *The emerging spatial mind* (pp. 320–361). New York: Oxford University Press.

Spencer, J. P., Smith, L. B., & Thelen, E. (2001). Tests of a dynamic systems account of the A-not-B error: The influence of prior experience on the spatial memory abilities of 2-year-olds. *Child Development, 72*, 1327–1346.

Spivey, M. (2007). *The continuity of mind*. New York: Oxford University Press.

Thelen, E., Schöner, G., Scheier, C., & Smith, L. B. (2001). The dynamics of embodiment: A field theory of infant perseverative reaching. *Behavioral & Brain Sciences, 24*, 1–86.

Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge, MA: MIT Press.

Thelen, E., & Smith, L. B. (2006). Dynamic systems theories. In W. Damon (Series Ed.) & R. M. Lerner (Vol. Ed.), *Handbook of child psychology: Vol. 1. Theoretical models of human development* (6th ed., pp. 258–312). New York: Wiley.

van der Maas, H. L. J., & Molenaar, P. C. M. (1992). Stagewise cognitive development: An application of catastrophe theory. *Psychological Review, 99*, 395–417.

Van Geert, P. (1998). A dynamic systems model of basic developmental mechanisms: Piaget, Vygotsky, and beyond. *Psychological Review, 105*, 634–677.

Waddington, C. H. (1954). The integration of gene-controlled processes and its bearing on evolution. *Proceedings of the 9th International Congress of Genetics, 9*, 232–245.

Wang, X.-J. (2001). Synaptic reverberation underlying mnemonic persistent activity. *Trends in Neuroscience, 24*, 455–463.

Wynn, K. (1992). Addition and subtraction by human infants. *Nature, 358*, 749–750.

Child Development, November/December 2008, Volume 79, Number 6, Pages 1648 – 1653

In Defense of Change Processes

Karen E. Adolph
New York University

Scott R. Robinson
University of Iowa

Nativist and constructivist approaches to the study of development share a common emphasis on characterizing beginning and end states in development. This focus has highlighted the question of preservation and transformation—whether core aspects of the adult end state are present in the earliest manifestations during infancy. In contrast, a developmental systems approach emphasizes the process of developmental change. This perspective eschews the notions of objective starting and ending points in a developmental progression and rejects the idea that any particular factor should enjoy a privileged status in explaining developmental change. Using examples from motor development and animal behavior, we show how a developmental systems framework can avoid the pitfalls of the long and contentious debate about continuity versus qualitative change.

Continuity and Qualitative Change

Every developmental researcher agrees that the primary goal of developmental psychology is to understand development. As described in Kagan (2008), a popular approach to the problem of understanding development is to characterize beginning and end states in a developmental progression and to determine whether important structural and functional features in the initial incarnation are present in the later, adultlike appearance of the skill (Carey, 1991; Pinker, 1984; Spelke & Kinzler, 2007). A variant of this "stable state" approach to development is to document multiple stages between beginning and end states and to determine critical similarities and differences among them (Piaget, 1954). The focus on origins, stages, and end states has led researchers from nativist and constructivist frameworks to engage in a long and furious debate about the preservation versus transformation of earlier structures and functions in later, more mature forms. For these researchers, the tension between continuity and qualitative change remains a central issue for understanding development.

In the domain of conceptual development, Kagan points out that arguments for continuity between critical aspects of infants' and adults' conceptual knowledge are based largely on inferences drawn from a relatively lean database: the accumulated duration of young infants' looking times to unexpected visual displays. Based on his discomfort in ascribing conceptual knowledge to young infants based on changes in looking times, Kagan suggests that the notion of qualitative changes in development remains viable. He is in good company: Other researchers have criticized looking times as a basis for inference about infants' conceptual abilities (Bogartz, Shinsky, & Schilling, 2000; Thomas & Gilmore, 2004) and have thus argued in defense of qualitative differences between infant and adult forms in conceptual development.

Ironically, regardless of whether researchers locate themselves on the continuity or discontinuity side of the debate, the focus on characterizing beginning and end states has resulted in a shortage of research on the process of developmental change. The literature is filled with exciting lo-and-behold studies of the earliest manifestations of infants' abilities. Journals and textbooks abound with before and after snapshots of children's remarkably inept behaviors at Time 1 and more mature behaviors at Time 2. However, developmental scientists know surprisingly little about the process of development. Even Piaget's stages—the best-known characterization of children's mental structures—are so exasperatingly bereft of detailed change processes that Klahr (1982) referred to assimilation and accommodation as the "Batman and Robin" of developmental psychology (p. 80). Missing are accounts of developmental change processes.

This work was supported by National Institute of Child Health and Human Development Grants R37-33486 and R01-42697 awarded to K.E.A and by R01-33862 awarded to S.R.R. We thank Peter Gordon for his stimulating comments.

Correspondence concerning this article should be addressed to Karen E. Adolph, Department of Psychology, New York University, 4 Washington Place, Room 410, New York, NY 10003. Electronic mail may be sent to karen.adolph@nyu.edu.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7906-0007

An alternative, developmental systems framework (Gottlieb, 2007; Oyama, Griffith, & Gray, 2001; Thelen & Smith, 1994) may prove more useful for meeting the goal of understanding development. A developmental systems view sidesteps the debate over continuity and qualitative change—at least as it is formulated with reference to beginning and end states. Periods of relative stability are considered only as benchmarks because stable periods at the ends of a progression may be pushed ever outward or shown to be unstable under less favorable conditions. Common features between earlier and later appearing skills are not accorded privileged developmental status because mature forms are the result of multiple factors, each of which must be at a particular level of readiness. Researchers accept as a matter of course that some structures and functions are preserved (more or less), and some are not.

Most important, a developmental systems framework is concerned with process rather than end products. The focus is shifted from characterizing children's abilities at Time 1 and Time 2 to characterizing the trajectory between the periods of relative stability and the underlying mechanisms of change that shape the developmental path. Arguments about continuity and discontinuity are more likely to arise around the shape of the developmental trajectory, rather than around the preservation of structural or functional features from one time to the next.

Kagan focused on cognitive development and the problem of drawing strong inferences about conceptual structures from small changes in infants' looking times. However, the ideas raised in the article have broad implications for understanding developmental change in other domains. We expand on Kagan's ideas, drawing on evidence from motor development and animal behavior to illustrate our points.

Beginning and End States

Although the theoretical positions of nativism, as exemplified by the work of Spelke and her colleagues (Spelke & Newport, 1998), and constructivism, as exemplified by the work of Piaget and as espoused by Kagan in his target article, are diametrically opposed in terms of the continuity/qualitative change debate, they share a common perspective in discussing development in terms of preservation and transformation. Perhaps the debate is a natural outgrowth of comparing infants to older children or adults: Comparison of performance at Time 1 and Time 2 naturally leads to questions about what remains unchanged and what differs between the two times.

Moreover, both nativist and constructivist traditions share a tendency to reify the beginning and end states in a developmental progression, especially when the ages are sampled from infancy and adulthood. The earliest times at which a particular experimental method can be applied are often interpreted as the starting point for development of the phenomenon, and infants' success in the task is interpreted as core knowledge (Spelke & Kinzler, 2007) or the primary repertoire (cf. Sporns & Edelman, 1993). An arms race has emerged in this game, with methodological advances allowing researchers to tout infants' abilities at earlier and earlier ages.

The assumption of an initial condition and a mature end state in development is not limited to conceptual knowledge, grammar, or other cognitive abilities. In motor development, where structure and form can be observed directly rather than inferred from outward behaviors such as looking times, many researchers also have assumed beginning and end states in development. Adult walking, for example, can be characterized by alternating movements of the limbs within a girdle (the two legs in humans and birds and the forelimbs and hindlimbs in quadrupeds) and by alternating flexor and extensor muscle activation within each limb. Long before infants can walk, however, they can move their legs in alternation. When newborn infants are held upright, they alternate their legs in slow marching movements; while lying on their backs, infants spontaneously kick their legs using many of the same alternating coordination patterns; and ultrasound recordings confirm alternating leg movements in the human fetus (Suzuki & Yamamuro, 1985; Thelen & Fisher, 1982). Before birth or hatching, the rat fetus and chick embryo display alternating leg movements during spontaneous and pharmacologically induced motor activity (Bekoff, 1992; Brumley & Robinson, 2005). Even more dramatic, both adult and infant animal preparations can generate alternating limb activity with spinal cord circuitry alone. Moreover, dedicated neural structures in the spinal cord—termed "central pattern generators" (CPGs)—have a distinct window of development. In the rat, CPGs begin to produce bursts of alternating activity in the spinal nerves that project to limb muscles 3–4 days before birth.

The presence of alternating coordination patterns in the legs and spinal cords of infants and fetuses has led many prominent researchers in motor development to argue that the foundations of locomotion are established prenatally, presumably when CPGs become active (Grillner & Wallen, 2004; Prechtl, 1997). Similarly, Spelke and Newport (1998) in their influential *Handbook* chapter interpreted the early coordination patterns as "inborn, foundational action capacities that persist over development and are common across a great range of animal species" and that "provide the core of humans' ability to locomote" (p. 278). According to this stable state view of motor development, the basic locomotor pattern may be masked at some points in development and later modulated by perceptual feedback and experience, but it remains unchanged in its fundamental form and organization (Grillner & Wallen, 2004).

Change Processes

Regardless of whether researchers lean toward nativism or constructivism, continuity or qualitative change, developmental psychology suffers from a lack of focus on change processes. One suggested solution is that researchers should take seriously the problem of describing the developmental path from Time 1 to Time 2 (Siegler, 2006; Vygotsky, 1978). How does infants' differential looking at an impossible physical event become elaborated or transformed into adults' understanding of physical constraints on objects? How does infants' alternating leg movements develop into running across a cluttered living room floor? A developmental systems approach urges researchers to embrace the challenge of explaining change processes and advocates that they take the idea one step further by rejecting the notion of a privileged factor that is core or foundational in engendering development and by rejecting the assumption of an initial starting point or a mature end state. According to the developmental systems perspective, there are no developmentally special Time 1s and Time 2s. Process is all there is. As stated so eloquently by Hutton (1795) when explaining the geological history of the earth, "we find no vestige of a beginning, no prospect of an end" (p. 80).

Starting and ending points are arbitrary. The rich motor, sensory, and learning abilities of the fetus should render suspect any claims about the primacy of birth in defining initial conditions in development. Other developmental benchmarks such as conception, puberty, and death are salient but not privileged starting or ending points in development. Environmental, organismic, physiological, and genetic factors can exert their influences beyond these boundaries.

Developmental systems researchers address the question of developmental origins by examining the multicausal, multidimensional, and temporally complex details of change processes. For example, it may seem reasonable to locate the developmental roots of sex-typed behavior (e.g., sexual receptivity, aggression) in puberty when the testes and ovaries produce hormones at elevated levels. However, experiments with rodents and other animals that bear multiple offspring indicate that the initial starting point for sex-typical behavior in adulthood can occur long before puberty, before birth, and even before conception. The roots of sex-typed behavior can be located in the rodent dam or in her mother or her mother before her. Female gerbils exhibit masculinized behaviors as adults (display more aggression, more mounting, less receptive lordosis, etc.) if they gestate in a prenatal environment predominated by male fetuses where they are exposed to higher levels of male androgens during sensitive periods of brain development (Ryan & Vandenbergh, 2002). They also exhibit masculinized hormonal development that biases them to produce male-dominated litters themselves, perpetuating the effects of prenatal masculinization across generations (Clark, Karpiuk, & Galef, 1993). Moreover, hormonal exposure is not a privileged factor in sex-typed development. Other effects on sex-typed behaviors can be nonobvious but equally important: Different amounts of maternal licking of rat pups' anogenital region (an important part of maternal care) can bias and accentuate tendencies to express male- or female-typed behaviors later in life (Moore, 1995).

Returning to the example of alternating leg movements in the development of walking, what are we to conclude from the compelling demonstrations of coordinated alternation in neonates and fetuses, the dedicated neural circuitry that seems to underlie the coordination pattern, and the striking similarity between the alternating coordination patterns in infants and adults? From a stable state perspective, the alternating pattern represents a core capacity in adult walking and reveals continuity between beginning and end states (Spelke & Newport, 1998). The developmental systems perspective offers a different interpretation: The activation of a CPG is not a privileged starting point for the development of locomotion, the alternating pattern is no more "core" than the multitude of factors that contribute to the development of walking, and there is no objective end point for adult gait.

The activation of a CPG is not a privileged starting point for the development of walking because alternating bursts of activity in the ventral roots of spinal nerves is not a discrete developmental event. As with all phenomena, the alternating pattern has a developmental history. In the rat fetus, days before spinal nerves show the alternating pattern, the same neural circuitry generates synchronized bursts of activity that result in spontaneous limb movements that are not alternating (Bekoff, 1992). The different patterns likely result from developmental changes in the inhibitory neurotransmitter systems. Gamma-aminobutyric acid (GABA) and glycine, the two most important inhibitory neurotransmitters in the central nervous system, are excitatory rather than inhibitory in the spinal cord of younger fetuses.

Moreover, after the alternating pattern becomes available, it is not obligatory. Human infants spontaneously and deliberately move their legs in a variety of coordination patterns while kicking, crawling, and upright throughout the 1st year. Coordination patterns are plastic and can be flexibly adapted to local conditions. For example, tying rat fetuses' hind legs together with a tiny elastic string shifts the spontaneous pattern of interlimb coordination from alternating to synchronous, and the new pattern is maintained after the string is cut (Robinson, 2005). A similar manipulation shifts the alternating pattern to a synchronous one in human infants (Thelen, 1994).

Alternating leg movements are not a privileged developmental factor in the acquisition of walking. The alternating pattern is only one of the many critical factors that must be in place for infants to walk. The list also includes balance control, muscle strength, appropriate body proportions, sensitivity to visual flow, the ability to detect affordances for upright locomotion, and the motivation to go somewhere (Thelen & Smith, 1994). Each of these factors shares some commonalities and some differences with adult walking, and each has a different developmental time course. For example, although alternation is a key aspect of adult gait in humans, rats, and cats, many mammals that display the alternating pattern during fetal and infancy periods do not alternate their legs as adults (Eilam, 1997). Hopping and bounding are the mature gait patterns of kangaroos, rabbits, and desert rodents such as gerbils, jerboas, and kangaroo rats. More important, the transition from alternation to synchronous hopping occurs postnatally, coincident with rapid changes in leg morphology. From the developmental systems perspective, the goal of developmental analysis is to understand the paths, interrelations, and causal mechanisms of the various time course of each contributing factor.

Defining an end point to development may be an equally difficult enterprise. Structures and functions assumed to be stable features of the adult skill may turn out to be stable only under certain rearing or testing conditions. The structure of adult walking, for example, has been characterized in exquisite detail in terms of inter- and intralimb timing, joint angles, the coordination of flexor and extensor muscle activation, and the dynamic forces produced at each point in the step cycle. However, different rearing and testing conditions can lead to very different end points: A family of adult siblings in Turkey crawl on their hands and feet rather than walk, although walking was physically possible (Humphrey, Skoyles, & Keynes, 2005). A dog born without forelegs developed smooth and functional bipedal locomotion (Stringfellow, 2007). A thousand years of foot binding in China resulted in billions of girls with altered gait patterns to accommodate balance and propulsion on three-inch long feet (Adolph, Karasik, & Tamis-LeMonda, in press). Women in western Kenya have altered the dynamic characteristics of their gait patterns so as to carry tremendous loads on their heads (up to 70% of body weight) with reduced energetic costs (Heglund, Willems, Penta, & Cavagna, 1995). Endurance running is so advanced in the Tarahumaran Indians of the American Southwest that they regularly run the distance of six modern marathons in 1–2 days time (Devine, 1985). Such individual and cultural variations in developmental end points are windows on developmental process and are not as readily accommodated by presumptions of species-typical developmental patterns leading to normal adult behavior.

Methodological Challenges and Conceptual Payoffs

As Kagan reminds us, conclusions about development are only as good as the data. Developmental research has suffered from overreliance on particular experimental paradigms, such as looking time in infants' conceptual development. Strong interpretations about children's knowledge or skills warrant rich data sets derived from convergent behavioral and physiological measures obtained in multiple paradigms. Robust conclusions about the process of change require robust sources of evidence, including adequate sampling to observe change as it occurs.

One strategy for choosing appropriate developmental measures is to use methods that can be applied across a range of ages. The problem with this strategy, as Kagan points out, is that the same measure may reflect different functions or underlying mechanisms at different ages. Kagan suggests that this problem can be avoided by measuring variables at different levels of analysis, presumably closer to the underlying neural and physiological mechanisms governing the response. Physiological measures, such as brain imaging, heart rate, skin conductance, and electromyographic recording, can be sampled across a broad range of ages.

An important caveat, however, is that measurement of variables at lower levels of analysis does not ensure relevance for understanding processes of developmental change. Changes in looking time or any other behavioral measure are accompanied by dramatic changes in body dimensions, physiology, neuroanatomy, neurochemistry, and the growth, interaction, and functioning of cells. These changes are parallel to and deeply interactive with changes in behavior; the direction of causal influence points up and down levels of analysis. Neural, motor, and sensory experiences play dramatic roles in shaping the course of brain development. A developmental systems perspective is particularly helpful from this standpoint because the emphasis on multicausality and multiple levels of analysis emphasizes the need for multiple convergent lines of evidence for characterizing any developing behavior.

Ultimately, we may find no substitutes, only supplements, for direct behavioral assessments, and so the selection of age-appropriate behavioral tasks will remain an important consideration in planning developmental research. In much of the current developmental research, behavioral measures have become more specific and restricted. Kagan points out that total looking time ignores the duration of looking bouts, the number of shifts of attention, the location of eye gaze, and other measures of visual exploration that can be captured from video and eye-tracking recordings. Highly simplified tasks such as change in total duration of looking time facilitate the ease of measuring infant responses but fail to take advantage of infants' rich repertoire of behaviors (e.g., facial expressions, reaching, postural and locomotor responses, social expressions and interactions) that are likely to prove informative when viewed as additional, convergent sources of evidence about infants' perceptions, emotions, and cognitive abilities. Overreliance on a single paradigm may present an overly restricted picture of the diversity of infants' abilities.

A consequence of shifting the focus from identifying beginning and end states to understanding processes of change is choosing appropriate measures and regimes for sampling behavior across development. Surprisingly, the issue of how to make an informed choice about the rate of sampling developmental change has received little attention in developmental psychology. The legacy of nativist and constructivist traditions is that infrequent sampling is viewed as sufficient to identify the major stages and transformations that occur in the course of development. But infrequent sampling can grossly distort patterns of variability and the overall shape of change in the path of development.

Using a data set of infant motor skills sampled daily during the first 18 months and resampled to simulate alternative regimes ranging from 2 to 31 days, we have reported that sampling at rates as frequent as weekly can seriously underestimate the amount of variability relative to daily sampling of motor skills (Adolph, Robinson, Young, & Gill-Alvarez, 2008). Sensitivity to variability is reduced as an inverse power function of the rate of sampling, such that skills that fluctuate between expression and nonexpression 10 or 20 times before achieving a period of stable occurrence erroneously look like single stage-like transformations when indexed weekly or monthly. In fact, 84% of the 261 data sets were characterized by variable trajectories when sampled daily, but 93% appeared to show a single stage-like appearance of the skill when sampled monthly. Estimates of the day of onset of a skill were similarly affected, with errors up to 109 days compared to daily sampling. If developmentalists are to take seriously the challenge of understanding processes of developmental change, then fresh looks at the shape of developmental change will require careful consideration of our methods and rates of collecting developmental data.

In summary, the developing child is a complex and interactive system. The way we view development therefore should transcend simple conceptions of preservation and transformation to characterize patterns and processes of change. We agree with Thelen and Smith (1998) that important progress can be made in understanding change processes if researchers abandon the old dualities such as continuity versus qualitative change and shift their focus to how the developing system works. The developmental systems approach holds additional promise. With a focus on describing the multilayered, multicausal path of development, researchers may gain a better understanding of underlying change mechanisms.

References

Adolph, K. E., Karasik, L. B., & Tamis-LeMonda, C. S. (in press). Moving between cultures: Cross-cultural research on motor development. In M. H. Bornstein (Ed.), *Handbook of cross-cultural development science: Vol. 1. Domains of development across cultures.* Mahwah, NJ: Erlbaum.

Adolph, K. E., Robinson, S. R., Young, J. W., & Gill-Alvarez, F. (2008). What is the shape of developmental change? *Psychological Review, 115,* 527–543.

Bekoff, A. (1992). Neuroethological approaches to the study of motor development in chicks: Achievements and challenges. *Journal of Neurobiology, 23,* 1486–1505.

Bogartz, R. S., Shinsky, J. L., & Schilling, T. H. (2000). Object permanence in five-and-a-half-month-old infants? *Infancy, 1,* 403–428.

Brumley, M. R., & Robinson, S. R. (2005). The serotonergic agonists quipazine, CGS-12066A and α-Methylserotonin alter motor activity and induce hindlimb stepping in the intact and spinal rat fetus. *Behavioral Neuroscience, 119,* 821–833.

Carey, S. (1991). Knowledge acquisition: Enrichment or conceptual change? In S. Carey & R. Gelman (Eds.), *The epigenesis of mind* (pp. 257–291). Hillsdale, NJ: Erlbaum.

Clark, M. M., Karpiuk, P., & Galef, B. G. (1993). Hormonally mediated inheritance of acquired characteristics of Mongolian gerbils. *Nature, 367,* 712.

Devine, J. (1985). The versatility of human locomotion. *American Anthropologist, 87,* 550–570.

Eilam, D. (1997). Postnatal development of body architecture and gain in several rodent species. *Journal of Experimental Biology, 200,* 1339–1350.

Gottlieb, G. (2007). Probabilistic epigenesis. *Developmental Science, 10,* 1–11.

Grillner, S., & Wallen, P. (2004). Innate versus learned movements—A false dichotomy? *Progress in Brain Research, 143,* 3–12.

Heglund, N. C., Willems, P. A., Penta, M., & Cavagna, G. A. (1995). Energy-saving gait mechanics with head-supported loads. *Nature, 375,* 52–54.

Humphrey, N., Skoyles, J. R., & Keynes, R. (2005). *Human hand-walkers: Five siblings who never stood up.* Retrieved May 1, 2008, from http://eprints.lse.ac.uk/archive/00000463

Hutton, J. (1795). *Theory of the earth: With proofs and illustration.* Edinburgh, Scotland: William Creech.

Kagan, J. (2008). In defense of qualitative changes in development. *Child Development, 79,* 1606–1624.

Klahr, D. (1982). Non-monotone assessment of monotone development: An information processing analysis. In S. Strauss & R. Stavy (Eds.), *U-shaped behavioral growth* (pp. 63–86). New York: Academic Press.

Moore, C. L. (1995). Maternal contributions to mammalian reproductive development and the divergence of males and females. In P. J. B. Slater, C. T. Rosenblatt, & M. Milinski (Eds.), *Advances in the study of behavior* (Vol. 24, pp. 47–118). New York: Academic Press.

Oyama, S., Griffith, P. E., & Gray, R. D. (2001). *Cycles of contingency: Developmental systems and evolution.* Cambridge, MA: MIT Press.

Piaget, J. (1954). *The construction of reality in the child.* New York: Basic Books.

Pinker, S. (1984). *Language learnability and language development.* Cambridge, MA: Harvard University Press.

Prechtl, H. F. R. (1997). State of the art of a new functional assessment of the young nervous system. An early predictor of cerebral palsy. *Early Human Development, 50,* 1–11.

Robinson, S. R. (2005). Conjugate limb coordination after experience with an interlimb yoke: Evidence for motor learning in the rat fetus. *Developmental Psychobiology, 47,* 328–344.

Ryan, B. C., & Vandenbergh, J. G. (2002). Intrauterine position effects. *Neuroscience and Biobehavioral Reviews, 26,* 665–678.

Siegler, R. S. (2006). Microgenetic analysis of learning. In D. Kuhn & R. S. Siegler (Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language* (6th ed., pp. 464–510). New York: Wiley.

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science, 10,* 89–96.

Spelke, E. S., & Newport, E. L. (1998). Nativism, empiricism, and the development of knowledge. In W. Damon (Ed.), *Handbook of child psychology: Vol. 1. Theoretical models of human development* (5th ed., pp. 275–340). New York: Wiley.

Sporns, O., & Edelman, G. M. (1993). Solving Bernstein's problem: A proposal for the development of coordinated movement by selection. *Child Development, 64,* 960–981.

Stringfellow, J. (2007). *Faith alone: Stories of an amazing dog.* Philadelphia: Xlibris.

Suzuki, S., & Yamamuro, T. (1985). Fetal movement and fetal presentation. *Early Human Development, 11,* 255–263.

Thelen, E. (1994). Three-month-old infants can learn task-specific patterns of interlimb coordination. *Psychological Science, 5,* 280–285.

Thelen, E., & Fisher, D. M. (1982). Newborn stepping: An explanation for a "disappearing reflex." *Developmental Psychology, 18,* 760–775.

Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action.* Cambridge, MA: MIT Press.

Thelen, E., & Smith, L. B. (1998). Dynamic systems theories. In W. Damon (Ed.), *Handbook of child psychology: Vol. 1. Theoretical models of human development* (5th ed., pp. 563–634). New York: Wiley.

Thomas, H., & Gilmore, R. O. (2004). Habituation assessment in infancy. *Psychological Methods, 9,* 70–92.

Vygotsky, L. S. (1978). *Mind in society: The development of higher mental processes.* Cambridge, MA: Harvard University Press.

Child Development, November/December 2008, Volume 79, Number 6, Pages 1654 – 1658

Methodological and Epistemological Issues in the Interpretation of Infant Cognitive Development

Ulrich Müller and Gerald Giesbrecht
University of Victoria

This commentary on J. Kagan (2008) addresses 2 issues. The first concerns the importance of studying developmental sequences and processes of change. The second concerns epistemological differences between contemporary neonativist approaches and classical theories of development. The commentary argues that classical theories of infant cognition and contemporary neonativist theories fundamentally differ in terms of how they conceptualize the workings of the mind. These differences affect the meaning of terms such as representation and reasoning. Also discussed are implications of epistemological differences for developmental explanations. The commentary concludes that epistemological differences must be articulated more fully in order to improve understanding of different theories and in order to evaluate the relative merits and shortcomings of those theories.

Kagan (2008) joins the chorus of recent critics who have argued that the rich interpretation of infant looking-time data is not warranted (e.g., Bremner & Mareschal, 2004; Cohen & Cashon, 2006; Newcombe, 2002; Rakison, 2007). Kagan challenges the rich interpretation of looking-time data on two grounds. First, he points out that the evidential basis for a rich interpretation is not as clear cut as one might wish. This is due to methodological problems with the interpretation of looking time data as well as the fact that conclusions about infant cognition rest on looking-time data only. To put the interpretation of findings on more solid ground, Kagan makes valuable suggestions regarding the use of a greater variety of measures in studies of infant cognition. The second challenge follows from the epistemological distinction between different knowing levels. Kagan distinguishes between a perceptual-sensorimotor level of knowing ("schemata") and a conceptual level of knowing ("semantic form"). He argues that findings generated in the looking-time (i.e., violation of expectation) paradigm do not require a rich interpretation but are consistent with a perceptual-level explanation. Kagan provides numerous excellent examples that we think convincingly drive this message home.

In our commentary, we address two issues. The first issue concerns the importance of studying developmental sequences and processes of change. The second issue concerns some epistemological differences between contemporary neonativist approaches and classical theories of development. We are skeptical that methodological refinements alone will settle the issue concerning the interpretation of findings from looking time experiments (and of infant cognition in general). This is because theories operate within different epistemological frameworks that make different assumptions about the workings of the mind. We will draw mostly on work by Baillargeon (on the neonativist side) and Piaget (on the classical side) to highlight these differences, but we could as well have used other theorists such as, on the one hand, Wynn and Leslie, and on the other hand, Baldwin, Stern, Werner, and Vygotsky.

The Importance of Studying Developmental Sequences and Processes of Change

An important point raised by Kagan (2008; see also Werner, 1957) is that similar behaviors and performances may be mediated by different cognitive operations or mechanisms. Despite similarities in performance over time, the underlying process by which outcomes are achieved may vary dramatically (Siegler, 1995), and these variations are indicative of similarities and differences in the cognitive abilities at different developmental levels. One way of disambiguating the cognitive basis of infants' performance is to place the performance at earlier developmental levels in the context of an extended developmental sequence and to interpret the initial steps in this sequence in the light of the last steps. Piaget terms

---

We would like to thank Jeremy Carpendale for helpful comments on an earlier draft of this article.

Correspondence concerning this article should be addressed to Ulrich Müller, Department of Psychology, University of Victoria, PO Box 3050 STN CSC, Victoria, BC, Canada V8W 3P5. Electronic mail may be sent to umueller@uvic.ca.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7906-0008

this the method of recurrence, which consists "in extending in the opposite direction the lines of the genetic process revealed by the study of the last stages. The ground for interpretations relating to the point of departure of a concept can only be shown a posteriori by the probability of the explanation based on the total evolution of that concept" (Piaget, 1937/1954, p. 221).

Studies of developmental sequences assist in determining whether a particular behavior at an earlier point in time is a precursor or prerequisite of behavior at a later point (Campbell & Richie, 1983). Furthermore, by designing studies of developmental sequences in a way that captures the processes involved in the transformation of abilities over time, constraints can be placed on the interpretation of the cognitive basis of behavior because understanding the mechanisms of change helps in determining the cognitive organization that characterizes behavior at different levels of development.

Epistemological Issues: It Is Not Just About What Is Innate Versus Learned

Neonativists (Baillargeon, 2008; Spelke, 1998) place their findings in the context of long-standing controversy between rationalism and empiricism. We are told that rationalist philosophers such as Descartes and Leibniz endorsed innate ideas, whereas empiricist philosophers such as Locke, Hume, and Mill rejected innate ideas and explained the acquisition of knowledge in terms of experience and domain-general processes operating on experience (Baillargeon, 2008). Contemporary findings are then claimed to shed light on this old epistemological controversy (Spelke, 1998). Leaving aside the issue that the epistemological controversy concerns the sources and techniques by means of which claims to knowledge are to be justified (which is a matter quite different from identifying the developmental origins of knowledge), a more careful reading of these philosophers suggests that their claims about innate ideas are more complex and cannot that easily be divided along the lines of nativism versus learning (Gorham, 2002; Greenlee, 1972; Saville, 1972). For example, Locke's theory of language does not refute innate ideas and is, in terms of theoretical assumptions, surprisingly similar to Pinker's theory of language (Proudfoot, in press). Overlooked in the fixation on the innate versus learning issue are deeper epistemological differences in terms of how the working of the mind is conceptualized in different theories. Next, we briefly illustrate some of these differences and show how they affect the meaning of terms such as representation and reasoning. Finally, we suggest that these epistemological differences have implications for (a) what different theories consider to be in need of explanation and (b) what form such explanations should take (Overton, 2006; Samuels, 2002).

Contemporary neonativism is based on a causal representational or mechanist theory of meaning, which itself is rooted in the empiricist tradition (Goldberg, 1991; McDonough, 1989). According to such theories, meaning and behavior are explained through internal representations that mediate between input and output. These representations are causally produced by input (e.g., perceptual information) and, in turn, effect some output (e.g., bodily movements). One's mental contents, or ideas, are just the effects of stimuli external to the mind and the mechanical operations or computations triggered by those stimuli. In this sense, these representations are not intentional, and there is no active agent producing them. Representations become entities that take on a life of their own. As a result, the "person as agent becomes superfluous" (Judge, 1985, p. 51).

Arguably, Baillargeon's (2008) account of infants' physical reasoning instantiates a causal (empiricist) representational theory of meaning. According to Baillargeon, infants are born with the principle of persistence. This principle holds that "objects persist, as they are, in time and space" (Baillargeon, 2008, p. 3). This principle is part of a physical reasoning system, which is characterized "as an abstract computational system designed to monitor events as they unfold and to interpret and predict their outcomes" (Baillargeon, 2008, p. 3). When infants observe a pertinent event, the physical reasoning system is activated and it builds a representation of the event. For example, "while watching a red ball being alternately lowered behind and lifted above a screen, infants would represent the information 'inert closed object being alternately lowered behind and lifted above inert closed object'" (Baillargeon, 2008, p. 4). The representation becomes enriched as, with experience, infants include—through a process of association—more and more information in the representation of the event.

Baillargeon's account of the development of physical reasoning exemplifies a causal representational theory of meaning because the representations of events are triggered by some input and run through some computations that are performed by the physical reasoning system, which then produces an output (e.g., increased attention to the physically impossible event). The representations are unintentional in the same way as, say, the patellar reflex: It is not the infant who is doing the reasoning but some device inside the infant's head computes the representation. The computation and representation are happening to the infant; the infant is herself passive. For that reason, it is somewhat a misnomer to speak in this instance of "infant reasoning" because these processes are subpersonal and performed by some device, as Baillargeon (2004, p. 422) explicitly states: "It seems very unlikely that infants possess explicit beliefs about anything. What they do possess is an abstract computational system, a physical reasoning system that monitors physical events and flags those that do not unfold as expected for further scrutiny." In this sense, these processes might also be likened to, for example, computational processes that are performed by the visual system to achieve perceptual constancies.

To highlight the importance of epistemological frameworks, let us contrast Baillargeon's empiricist theory with Piaget's constructivist sensorimotor theory. Somewhat ironically, Baillargeon (2008) classifies Piaget as an empiricist, even though Piaget's critical stance toward empiricist theories should be well known (Piaget & Inhelder, 1969). For Piaget (1936/1963), the basic functions with which the infant interacts with the world are assimilation and accommodation. Piaget's notion of assimilation, at the psychological level, particularly highlights the active, intentional nature of the mind. Specifically, assimilatory schemes (e.g., grasping, pushing) comprise a need and are directed toward specific goals; that is, they refer to the particular elements or objects toward which human activity, and consciousness more specifically, is directed. By assimilating objects to preexisting behavioral or cognitive schemes, the individual confers some kind of meaning on these, whether it be simply of a functional sort or (later on) more abstract in nature. Accommodation, on the other hand, is the aspect of the activity that provides the particular content for assimilation. Piaget holds that by accommodating to the environment, an organism's existing schemes are differentiated and modified. To illustrate, by assimilating different objects to the grasping scheme, these objects acquire the functional meaning "graspable" while at the same time, the infant learns to adjust her grasp to the specifics of different objects (accommodation). Meaning is created by the infant casting a web of intentionality onto the world. In fact, Piaget's action-based approach to knowledge acquisition and the working of the mind was designed to overcome the problems inherent in empiricist causal representational theories of mind such as Baillargeon's (see Bickhard, 1999; Bickhard & Terveen, 1995).

Even our necessarily cursory discussion of this matter should make it clear that there are fundamental differences in the way in which contemporary neonativist theories (exemplified by Baillargeon's theory) and classical approaches (exemplified here by Piaget's theory) conceptualize the working of the mind. Basically, in the tradition of empiricist causal representational theories, Baillargeon's theory conceptualizes the mind as passive and nonintentional, whereas Piaget's theory conceptualizes the mind as active and intentional (for a more detailed treatment, see Müller & Newman, 2008; Müller, Sokol, & Overton, 1998).

It is not surprising that, given these different epistemological frameworks, the same terms take on different meanings in Baillargeon's and Piaget's theories. Whereas for Baillargeon, representation is the effect of some input, this cannot be the case for Piaget. Furthermore, Piaget (1936/1963, pp. 242–243) distinguishes between two different meanings of representation. In one meaning, Piaget (1936/1963, p. 242) refers to representation as nonsymbolic meaning. This type of representation is evident in "the capacity to confer upon things a meaning before the action which this meaning permits. . . . The act of looking under a shawl for a shoe in order to strike it with a piece of wood (Obs. 129) is the prototype of this behavior pattern." Representation in the sense of nonsymbolic meaning is thus evident in anticipation, but nonsymbolic meaning can be traced back to the very beginning of mental life (Piaget, 1936/1963, p. 38).

In the second meaning, representation refers to the capacity to use symbols and signs to evoke an absent object. To our knowledge, neonativist have not yet spelled out whether the representations their experiments unveil are located at the level of nonsymbolic meanings or symbolic-representational meanings. Part of the reason for this is, we suspect, that Piaget's and neonativist theories operate with fundamentally different concepts of representation, whereas within Piaget's theory, it makes sense to distinguish between nonsymbolic and symbolic-representational meaning, this distinction does not make much sense when representation is conceived as an operation of a computational device.

Reasoning is another term that takes on a very different meaning in Baillargeon's and Piaget's theories. As we have seen, for Baillargeon (2008) reasoning is a subpersonal process that runs off in a computational device. Piaget (1936/1963, pp. 239–240, 267–268), too, wrote about reasoning in infancy. Specifically, he drew a functional analogy between particular sensorimotor action patterns that subordinate one scheme as means (e.g., pushing away an obstacle) to another scheme (e.g., grasping an object) as end and the subordination of premises to conclusions at the level of verbal intelligence.

Piaget thus ascribes reasoning to infants when their actions put objects into a hierarchical relation. For Piaget, sensorimotor "reasoning" is thus not a subpersonal process but rather indicated by a certain level of complexity of infants' actions in pursuit of their goals. Of course, it should be noted that for Piaget (1936/1963, p. 240; 1947/1950, pp. 220–223), sensorimotor reasoning is qualitatively different from reasoning on the symbolic, reflective plane. In fact, Piaget uses some of the same criteria to distinguish practical sensorimotor intelligence from reflective, symbolic intelligence as Kagan does for demarcating schematic knowledge and semantic-conceptual knowledge.

Epistemological frameworks have implications for the types of things that different theories consider to be in need of explanation. In Baillargeon's theory, the infant is essentially a spectator, or "couch potato," as Willatts (1997) termed it, who takes a reflective, contemplative attitude to the events she is presented with, reasons about them, and searches for explanations of unexpected events (Baillargeon, 2008). What Baillargeon's (2008) theory aims to explain, then, is how infants initially impoverished representations become gradually richer. Baillargeon's view of the infant differs markedly from that of classical theories, which argue that the contemplative attitude is secondary and develops out of a pragmatic, unreflective attitude. Objects are originally perceived as *things of action* (Werner & Kaplan, 1963): They are imbued with affect, closely tied to the current situation and the satisfaction of immediate needs. As a result, classical theories of development aim at explaining how the infant gradually comes to distance herself and detach meaning from the immediate situation such that the infant is no longer dominated by immediate impressions and impulses (Piaget, 1936/1963; Vygotsky, 1933/1978; Werner & Kaplan, 1963). This notion of distancing is nicely captured in the following remarks by Gehlen (1940/1988):

> The things in our world thus have the essentially human character of *acquired neutrality*. This is not the indifference toward anything that does not appeal directly to instincts, as is the case in the environments of the higher animals. In contrast, the things around us are thoroughly known and "worked through," but remain "undecided" for the most part, available for interaction at any time. This is how man masters the overwhelming barrage of impressions, this is how he obtains relief: He *actively* "checkmates" *the forcefulness of the world's impressions, making the world potentially available at any time.* (pp. 162–163, emphases in original)

It is difficult to see how the notion of distancing could have a place in Baillargeon's theory. Given that representations are computed and interpreted by a physical reasoning system, there is no place for a distancing process. The objectivity of the world is built into the modular device and, as such, vanishes.

Finally, it should come as no surprise that epistemological differences affect the form that explanations take in different theories (for a detailed discussion, see Overton, 2006). Although in Baillargeon's (2008) theory, explanations invoke quantitative changes in the efficiency of computations and the type of information that is processed, classical theories (e.g., Piaget, 1936/1963; Werner, 1957) invoke processes such as differentiation and integration as explanations of the distancing processes.

Conclusions

Kagan makes numerous excellent points and provides several excellent suggestions that, if heeded, will undoubtedly help provide the field of infancy research with a better evidential basis. However, we are skeptical that the controversy over the interpretations of empirical findings on infant cognition will be settled by more methodological sophistication alone. In fact, it is not clear to us that the distinction between different knowing levels suggested by Kagan could even be meaningfully applied to the computational devices outlined in neonativist theories. Furthermore, deeper lying epistemological differences color the meaning of the terms being used, the goal of the explanations, and the very types of explanations thought. We believe that more effort must be put into articulating these epistemological differences. Only in this way will it be apparent that terms (e.g., representation) have meanings in one theory that are very different from their meanings in another theory. Eventually, such effort will also help us to better understand and evaluate the merits and shortcomings of different theories.

References

Baillargeon, R. (2004). Can 12 large clowns fit in a Mini Cooper? Or when are beliefs and reasoning explicit and conscious? *Developmental Science, 7*, 422–424.

Baillargeon, R. (2008). Innate ideas revisited: For a principle of persistence in infants' physical reasoning. *Perspectives on Psychological Science, 3*, 2–12.

Bickhard, M. H. (1999). Interaction and representation. *Theory and Psychology, 9,* 435–458.

Bickhard, M. H., & Terveen, L. (1995). *Foundation issues in artificial intelligence and cognitive science: Impasse and solution.* Amsterdam, Netherlands: Elsevier.

Bremner, A. J., & Mareschal, D. (2004). Reasoning . . . what reasoning? *Developmental Science, 7,* 419–421.

Campbell, R. L., & Richie, D. M. (1983). Problems in the theory of developmental sequences: Prerequisites and precursors. *Human Development, 26,* 156–172.

Cohen, L. B., & Cashon, K. H. (2006). Infant cognition. In W. Damon & R. Lerner (Series Eds.) & D. Kuhn & R. Siegler (Vol. Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language* (6th ed., pp. 214–251). New York: Wiley.

Gehlen, A. (1988). *Man, his nature and place in the world.* New York: Columbia University Press. (Original work published 1940)

Goldberg, B. (1991). Mechanism and meaning. In J. Hyman (Ed.), *Investigating psychology* (pp. 48–66). London: Routledge.

Gorham, G. (2002). Descartes on the innateness of all ideas. *Canadian Journal of Philosophy, 32,* 355–388.

Greenlee, D. (1972). Locke and the controversy over innate ideas. *Journal of the History of Ideas, 33,* 251–264.

Judge, B. (1985). *Thinking about things: A philosophical study of representation.* Edinburgh, Scotland: Scottish Academic Press.

Kagan, J. (2008). In defense of qualitative changes in development. *Child Development, 79,* 1606–1624.

McDonough, R. (1989). Towards a non-mechanistic theory of meaning. *Mind, 98,* 1–21.

Müller, U., & Newman, J. L. (2008). The body in action: Perspectives on embodiment and consciousness. In W. F. Overton, U. Müller, & J. L. Newman (Eds.), *Developmental perspectives on embodiment and consciousness* (pp. 313–341). New York: Taylor Francis.

Müller, U., Sokol, B., & Overton, W. F. (1998). Reframing a constructivist model of the development of mental representation: The role of higher-order operations. *Developmental Review, 18,* 155–201.

Newcombe, N. S. (2002). The nativist-empiricist controversy in the context of recent research on spatial and quantitative development. *Psychological Science, 13,* 395–401.

Overton, W. F. (2006). Developmental psychology: Philosophy, concepts, methodology. In W. Damon & R. Lerner (Eds.), *Handbook of child psychology: Vol. 1. Theoretical models of human development* (6th ed., pp. 18–88). New York: Wiley.

Piaget, J. (1950). *The psychology of intelligence.* Totowa, NJ: Littlefield. (Original work published in 1947)

Piaget, J. (1954). *The construction of reality in the child.* New York: Basic Books. (Original work published in 1937)

Piaget, J. (1963). *The origins of intelligence in children.* New York: Norton. (Original work published in 1936)

Piaget, J., & Inhelder, B. (1969). The gaps in empiricism. In A. Koestler & J. Smythies (Eds.), *Beyond reductionism* (pp. 118–159). Boston: Beacon.

Proudfoot, D. (in press). Meaning and mind: Wittgenstein's relevance for the "Does language shape thought?" debate. *New Ideas in Psychology.*

Rakison, D. H. (2007). Is consciousness in its infancy in infancy? *Journal of Consciousness Studies, 14,* 66–89.

Samuels, R. (2002). Nativism in cognitive science. *Mind & Language, 17,* 233–265.

Saville, A. (1972). Leibniz' contribution to the theory of innate ideas. *Philosophy, 47,* 113–124.

Siegler, R. S. (1995). How does change occur: A microgenetic study of number. *Cognitive Psychology, 28,* 225–273.

Spelke, E. S. (1998). Nativism, empiricism, and the origins of knowledge. *Infant Behavior & Development, 21,* 181–200.

Vygotsky, L. S. (1978). The role of play in development. InL. S. Vygotsky (Ed.), *Mind in society* (pp. 92–104). Cambridge, MA: Harvard University Press. (Original work published in 1933)

Werner, H. (1957). The concept of development from a comparative and organismic point of view. In D. B. Harris (Ed.), *The concept of development: An issue in the study of human behavior* (pp. 125–148). Minneapolis: University of Minnesota Press.

Werner, H., & Kaplan, B. (1963). *Symbol formation.* New York: Wiley.

Willatts, P. (1997). Beyond the "couch potato" infant: How infants use their knowledge to regulate action, solve problems, and achieve goals. In G. Bremner & A. Slater (Eds.), *Infant development: Recent advances* (pp. 109–135). Hove, UK: Psychology Press.

Child Development, November/December 2008, Volume 79, Number 6, Pages 1659–1675

Automatic Activation of Adolescents' Peer-Relational Schemas: Evidence From Priming With Facial Identity

Lauri Nummenmaa
MRC Cognition and Brain Sciences Unit and University of Tampere

Kätlin Peets
University of Turku

Christina Salmivalli
University of Turku and University of Stavanger

This study provides experimental evidence for automatic, relationship-specific social information processing in 13-year-old adolescents. Photographs of participants' liked, disliked, and unknown peers were used as primes in an affective priming task with happy and angry facial expression probes and in a hypothetical vignette task. For the affective priming, reaction times were faster for congruent than for incongruent prime–probe pairs when the prime visibility was high and the prime–probe stimulus onset asynchrony was long. In the vignette task, participants attributed more hostility toward the protagonist, experienced more anger, and were more likely to retaliate when the disliked peer served as a prime. It is concluded that peer-relational schemas and related affect are activated automatically upon perception of a peer.

The social information processing (SIP) model (Dodge, 1986; reformulated by Crick & Dodge, 1994), has been one of the most influential theoretical models used to explain the behaviors of children and adolescents. According to the model, children enter social situations with relatively stable knowledge structures, such as social schemas, that guide online processing of situational cues. The online processing involves encoding and interpretation of cues, clarification of goals for the situation, response access, and finally, response decision. The present study was designed to address three issues that have received extensive attention in the literature but have rarely been studied empirically: (a) automaticity, (b) relationship specificity, and (c) emotional aspects of SIP.

The SIP model has been often used as a heuristic to explain the development of adversities in the peer group, such as aggression and rejection by other peers. For instance, several studies have shown that aggressive and rejected children are inclined to interpret intentions of others as hostile (e.g., Dodge, 1980; Dodge, Murphy, & Buchsbaum, 1984) and generate aggressive problem-solving strategies (Dorsch & Keane, 1994). Most of SIP is proposed to take place in an automatic fashion; that is, it cannot be inhibited. However, although Crick and Dodge (1994, p. 79) acknowledged the importance of incorporating "techniques for assessing automatic processes (using measures such as response time and assessment of priming effects)" to study social behaviors, research testing the SIP model has mainly relied on reflective measures and has thus not tested the automaticity assumption (Orobio de Castro, 2004; Orobio de Castro, Veerman, Koops, Bosch, & Monshouwer, 2002). Usually, hypothetical vignettes are presented to children, followed by questions assessing different SIP steps. Although some studies have aimed at differentiating automatic versus more reflective information processing by instructing children to respond immediately or after waiting for a while prior to giving the response (Orobio de Castro, Bosch, Veerman, & Koops, 2003; Rabiner, Lenhart, & Lochman, 1990), studies actually tapping online processing are practically nonexistent.

Second, the authors of the SIP model stressed the importance of considering social context when studying social cognitions (see Crick & Dodge, 1994). For instance, Feldman and Dodge (1987) found that differences between low- and high-status children with regard to SIP were evident only in a teasing situation,

---

This research was supported by the Academy of Finland Grants 119088 to L.N. and 202554/68884 to C.S. We thank the participating children, their parents, and teachers for making the research possible; Pasi Kainulainen for his help in preparing the materials and gathering the data; and Andrew Engell for his helpful comments on an earlier version of this article.

Correspondence concerning this article should be sent to Lauri Nummenmaa, Department of Psychology, University of Tampere, FIN-33014 Finland. Electronic mail may be sent to lauri.nummenmaa@uta.fi.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7906-0009

providing support for the Person × Situation approach. However, with some exceptions, most of the earlier studies did not explicitly focus on studying contextual effects. More recent studies have demonstrated the role of the dyadic context in social-cognitive evaluations and behaviors (e.g., Card & Hodges, 2007; Hubbard, Dodge, Cillessen, Coie, & Schwartz, 2001; Ray & Cohen, 1997). For instance, Hubbard et al. (2001) showed that boys in mutually aggressive dyads attributed more hostility to each other than boys in randomly selected dyads. Burgess, Wojslawowicz, Rubin, Rose-Krasnor, and Booth-LaForce (2006) found that SIP was different in the friendship context as compared with other interaction partners. A conceivable explanation is that over time children have not only developed a general representation of their age-mates (Rudolph, Hammen, & Burge, 1995; Salmivalli, Ojanen, Haanpää, & Peets, 2005) but also representations specific to each peer and relationship.

The notion of relationship specificity is captured by mental structures called relational schemas, which "include images of self and others, along with a script for an expected pattern of interaction, derived through generalization from repeated similar interpersonal experiences" (Baldwin, 1992, p. 462). When children interact with different peers, peer-relational schemas are likely to be activated, influencing SIP. Moreover, not only social cognitions but also behaviors vary across different dyadic contexts. For instance, Card and Hodges (2007) demonstrated that although victimization occurs within friendship dyads, it is not as common as victimization in relationships characterized by mutual dislike (6.8% vs. 35%, respectively). Clear relationship effects on social cognitions can be found in middle childhood and adolescence: Social information is processed differently depending on the relationship partner (Peets, Hodges, Kikas, & Salmivalli, 2007). This supports the notion (Harter, Waters, & Whitesell, 1998) that the general view of peers develops during childhood and gets more fine-grained with age and experience. In contrast, younger children are not yet sensitive to different contexts. For instance, McDowell, Parke, and Spitzer (2002) found no context (i.e., whether they were with their mother, father, or a peer) effects on kindergarten children's goals and strategies.

Belongingness to the peer group also becomes increasingly important during middle childhood and early adolescence. Rejection by peers is associated with adjustment problems (for a review, see Parker & Asher, 1987), and it tends to be extremely stable (e.g., Coie, 1990), even when the child's behaviors change. The rejected status of a child is likely to be maintained through social-cognitive processes that are congruent with the affect felt toward him or her. For instance, the child's positive behavior might be attributed to unstable causes and negative behavior to stable dispositions (Hymel, 1986). Even if the original reasons for disliking the child (such as aggressiveness) do not exist any more, representations of the child and affect felt toward him or her by other peers might persist. It is likely that such affect-congruent social cognitions become automatized over the course of time and increase the likelihood of maintaining the rejected status of a child.

Third, although Crick and Dodge (1994) regarded emotions as an integral part of SIP, Lemerise and Arsenio (2000) integrated emotional processes into the SIP model more explicitly and thoroughly by suggesting that both experienced and perceived emotions influence interpretation of social information. This argument is supported by empirical findings. Negative emotional feelings induce or exacerbate attribution of hostility (Dodge & Somberg, 1987; Orobio de Castro, Slot, Bosch, Koops, & Veerman, 2003) and increase the likelihood of making more stereotypic judgments and relying on simple heuristic cues (Bodenhausen, Sheppard, & Kramer, 1994). Furthermore, others' emotional expressions influence SIP. For example, children attribute more hostile intent to angry provocateurs than to sad or happy provocateurs in hypothetical social situations (Lemerise, Gregory, & Fredstrom, 2005). Lemerise and Arsenio also suggested that the affective nature of the relationship with a peer (e.g., pleasure associated with a liked peer vs. anger associated with a hostile peer) influences SIP. For instance, children interpret the actions of their peers they personally like and dislike differently. Liked peers are perceived as less hostile than disliked peers (Peets, Hodges, & Salmivalli, 2008), and they are also held less responsible for their negative behaviors (Hymel, 1986). As relational schemas include cognitive as well as affective components (Baldwin, 1992), it could be expected that the activation of a particular relational schema triggers not only specific cognitions but also relationship-congruent affect. Moreover, when children have relationship histories with their peers, and especially when strong negative emotions are present (e.g., in the case of disliked peers), children tend to rely heavily on already established representations of these peers rather than on situational cues (e.g., Crick & Dodge, 1994). Given that emotional processes influence physiology and behavior automatically and rapidly (within hundreds of ms; Adolphs, 2002), it is surprising that experimental studies tapping relationship-specific affective processing in children and adolescents are nearly nonexistent.

Taken together, there is evidence that relationship-specific schemas of peers exist by early adolescence. There has been systematic interest in the role of such schemas in the processing of social information (e.g., Dodge & Rabiner, 2004), their automatic activation (Crick & Dodge, 1994), and their affective nature (Lemerise & Arsenio, 2000), but these phenomena have rarely been examined empirically. However, in the priming literature, automatic processing of affective and social information has been widely studied in adults. In typical affective priming experiments, participants' task is to categorize the affective valence (pleasant vs. unpleasant) of a probe stimulus preceded with a prime stimulus with congruent or incongruent valence. The valence of a probe stimulus is recognized faster if it is preceded by a prime stimulus of the same valence (for a review, see Klauer & Musch, 2003). The phenomenon has been demonstrated using a variety of primes, such as affective visual scenes (Hermans, De Houwer, & Eelen, 1994, Experiment 1) and facial expressions (Carroll & Young, 2005), and it is assumed to be highly automatic. Accordingly, we reasoned that a similar paradigm would be advantageous for studying automatic activation of adolescents' affective peer-relational schemas.

Three lines of evidence from studies with adults suggest that *facial identity primes* would activate affective, relationship-specific schemas in adolescents. First, influential models of face recognition (Bruce & Young, 1986; Haxby, Hoffman, & Gobbini, 2000) posit that when the face of a specific person is perceived, the person- and identity-related information is automatically retrieved from long-term memory to facilitate social interaction with the person. Second, functional brain imaging studies have shown that viewing faces of close friends and family members is associated with modulation of neural responses in areas which mediate episodic memory, representation of mental states of others, and emotional responses (Gobbini, Leibenluft, Santiago, & Haxby, 2004; Leibenluft, Gobbini, Harrison, & Haxby, 2004). As relational schemas consist of information about relationship-specific cognitions and emotions stored in long-term memory, the aforementioned data can also be interpreted as activation of relational schemas. Third, it has been argued that perception of names or faces of personally significant versus unimportant persons results in relationship-specific priming effects. In a study by Baldwin, Carrell, and Lopez (1990, Experiment 2), Catholic and non-Catholic students first read a sexual passage after which they were primed with an angry picture of either an unknown professor or the Pope and subsequently filled in a self-concept questionnaire. The Catholic participants who were primed with the Pope rated their self-concept as lower than those primed with the professor, thereby suggesting a relationship-specific priming effect. Similarly, Banse (1999) demonstrated that masked names and pictures of friends and romantic partners led to a more positive evaluation of subsequently presented Chinese ideographs. In a follow-up study, Banse (2001) showed that a brief presentation of the names and faces of the participants' relationship partner and Saddam Hussein (Experiment 1) or Charlie Chaplin and Hussein (Experiment 2) sped up the categorization of probe words with congruent (pleasant vs. unpleasant) affective valence. Importantly, the prime faces were intrinsically neutral as they all had neutral facial expressions. However, the participants had associated the intrinsically neutral faces of Chaplin and Hussein with pleasant and unpleasant affective reactions, respectively, via learning. Consequently, Banse (1999, 2001) argued that presentation of the pictures elicited respective affective reactions that produced the priming effects.

Taken together, there is evidence to suggest that the perception of a face of a familiar (liked or disliked) person with a neutral facial expression can elicit relationship-specific affective and cognitive changes—at least in adult participants. However, as participants of the aforementioned studies were probably not personally familiar with the persons depicted in the prime stimuli (i.e., the Pope, Hussein, or Chaplin), it can be questioned whether the stimuli used by Baldwin et al. (1990) and Banse (2001) actually primed *relational* schemas. It is indeed likely that none of the participants had ever personally met, for instance, Chaplin or Hussein, but still they would evaluate the picture of Chaplin as more pleasant. This raises the question whether such primes can be implemented to study relational schemas or whether they just tap the reputation-based knowledge of the prime stimuli.

The Current Study

In the current study, we employed the priming paradigm to assess whether adolescents' peer-relational schemas would be automatically activated when perceiving a familiar peer and whether there would be an effect on subsequent SIP. We selected pictures of children's peers to serve as primes for two reasons. First, peers can be considered highly salient figures for individuals during middle childhood and adolescence (e.g., Larson & Verma, 1999).

Second, unlike names, faces provide unambiguous and rapidly accessible information about identity (Haxby et al., 2000). We chose adolescents as our participants because previous studies have found that early adolescents discriminate between different relationship contexts (e.g., Peets et al., 2007; Peets et al., 2008), whereas younger children do not (McDowell et al., 2002). In addition, participants of the present study had interacted with most of their classmates for more than 5 years. Thus, they had spent enough time interacting with different peers to form relationship-specific representations of them.

Our study implemented four methodological advances over the previous research on relationship priming. First, to assure that participants were familiar with both "unpleasant" and "pleasant" primes, we used pictures (digital photographs) of participants' disliked and liked classmates as primes. Second, we wanted to ensure that we were actually priming *peer-relational* schemas. Therefore, instead of using primes that might be classified as pleasant or unpleasant by *all* participants (e.g., using the "most liked" and "most disliked" children as primes), we selected unpleasant and pleasant primes by means of individual nominations for disliking and liking. Third, by manipulating the prime display duration and prime – probe stimulus onset asynchrony (SOA), we aimed at determining whether the relationship-specific affective information is extracted from faces pre- or postattentively. Preattentive affective processing (e.g., Bargh, 1997) would involve priming effects with short (< 50 ms) masked presentation (Baldwin et al., 1990) that does not allow detailed semantic or perceptual identification of the stimulus, whereas postattentive processing (Storbeck, Robinson, & McCourt, 2006) would involve priming effects with longer stimulus presentations (> 150 ms) and SOAs (> 300 ms; Calvo & Nummenmaa, 2007; Hermans, De Houwer, & Eelen, 2001). In Experiments 1 and 2, we used happy (congruent with liked primes) and angry (congruent with disliked primes) facial expression probes. We chose angry facial expressions, instead of other negative expressions such as fear, to represent probes congruent with disliked primes as our previously collected, unpublished data indicate that children have greater expectations of anger toward disliked than liked or neutral peers. If an affective component of peer-relational schemas was automatically activated upon face perception, faster choice reaction times (RTs) for congruent (i.e., disliked peer/angry expression and liked peer/happy expression) than incongruent (i.e., disliked peer/happy expression and liked peer/angry expression) prime – probe pairs could be expected.

Fourth, we wanted to assess whether the affective priming effects would be also evident with regard to SIP. As a complementary methodological approach, pictures of participants' liked and disliked peers were presented as primes before hypothetical provocation and rebuff vignettes. The vignettes were followed by questions tapping three aspects of SIP: attributions of hostility, experience of anger, and likelihood of retaliation. As Peets et al. (2007) demonstrated that hostile attributions and behavioral strategies were highly relationship specific by using a reflective measure, we were interested in examining whether similar target-specific cognitions would be triggered automatically. On the basis of the assumption that liked and disliked primes activate respective relational schemas automatically, we expected that when displaying disliked peers as primes, adolescents would attribute more hostility, experience more anger, and would be more likely to retaliate toward a hypothetical peer.

Experiment 1

In Experiment 1, participants' task was to categorize facial expressions (angry/happy) of probe pictures. Before the probe displays, pictures of faces of participants' disliked and liked classmates were presented. Pictures of same-age peers whom participants did not know were used as neutral primes. This resulted in incongruent (disliked/happy and liked/angry), neutral (neutral/angry and neutral/happy), and congruent (disliked/angry and liked/happy) prime – probe pairs. To manipulate the information available for visual processing of the pictures and to test whether relational schemas are activated by preattentive processing of the primes, we used brief (50 ms) structurally masked (low visibility) or unmasked (high visibility) presentations (see Breitmeyer & Ogmen, 2000) of the primes and assessed both costs and benefits resulting from prime – probe pairs with congruent and incongruent affective valence.

Method

Participants

Twenty-four 6th-grade children (12 girls, mean age = 13 years) participated in the experiment. The participants were randomly selected from a larger sample of 238 children (109 girls, mean age = 13 years) who were involved in a longitudinal study on peer relations in Turku, a midsized town (175,000 inhabitants) in southwest Finland. As the Finnish population is ethnically and socioeconomically diverse rather homogeneous, information about family background was not asked. In Turku, the households with school-aged children consist of entrepreneurs (4.6%), people in upper white-collar positions (20%), lower white-collar positions (27.6%), blue-collar positions (32.3%), and others (16.9%). About 94% of people in the area are of Finnish origin. All the participating children received parental permission to take part in the experiments. We examined whether the participants in Experiment 1 differed from the rest of the subjects who did not participate in any of the three experiments ($N = 180$) on peer-nominated social (rejection) and behavioral (withdrawal, aggression) reputation scores. No significant differences were found ($t \leq 1.80$). In addition, 216 children (108 girls, mean age = 13 years) recruited from among the adolescents involved in the longitudinal study on peer relations posed for the stimulus photographs used in the three experiments. All participants were compensated with a movie ticket.

Apparatus and Stimuli

The stimuli were presented with a 1.2 Ghz Toshiba Portégé computer on a 12-in. screen. Presentation computer software (Neurobehavioral Systems) controlled stimulus presentation and response collection. Response accuracy and latency were collected through key presses with a Cedrus RB-834 response pad. The initial fixation marker was a white cross (diameter 1°). The prime pictures were black-and-white photographs representing the classmates whom participants had nominated as their liked and disliked peers or photographs of unknown same-age peers. Two pictures from each category were used for each participant. The probe stimuli consisted of 28 black-and-white photographs of angry (14) and happy (14) facial expressions posed by different actors (7 male and 7 female) selected from a well-known standardized stimulus set of facial expressions (Ekman & Friesen, 1976). Both the prime and the probe displays measured 7 × 16° of visual angle and were presented at the center of the screen. All stimuli were presented against a black background.

Procedure

Identifying peers used as primes. For each participant, liked and disliked peers were identified by means of a standard sociometric procedure (e.g., Coie, Dodge, & Coppotelli, 1982) approximately 1 month prior to the experiment. Participants were given a sheet with all the names of their same-sex classmates and were asked to nominate up to three same-sex peers they liked the most and the least. Self-nominations were not allowed. For each participant, two nominated disliked and two liked peers were chosen as primes. All the adolescents who were selected to participate in the experiments had nominated at least two peers they liked and two peers they disliked. If a participant had nominated three liked or disliked peers, we selected the two prime peers randomly. In addition, for each participant, pictures of two same-age, same-sex peers outside the adolescent's own school (from a different part of the town) were randomly selected to serve as neutral (unknown) primes.

Prime pictures. The prime pictures depicting children with neutral facial expressions were photographed approximately 1 month before the experiments. The photographs were taken at a distance of 1 m, and a special viewfinder was used to center children's nose tips at the center of the image. At least three photographs were taken of each child and four independent judges selected the photograph with the most neutral expression to be later used in the experiment. Finally, the selected photographs were converted to grayscale and cropped to include only the face area of the posers and to conform to the size (7 × 16°) of the probe pictures.

Controlling for the facial attractiveness of the primes. Studies have consistently demonstrated that children perceive attractive peers as more likable and less mean than unattractive peers (e.g., Dion, 1973; Langlois & Cookie, 1978). Thus, any potential priming effect in this study could also be attributed to facial attractiveness instead of relational schemas. To control for this possible confound, a separate sample of 55 adolescents (mean age 13 years, 29 girls) rated the facial attractiveness of the prime pictures used in the three experiments using a scale ranging from 1 (*unattractive*) to 5 (*attractive*). Ratings were conducted with paper-and-pencil questionnaires that contained 3 × 4 cm black-and-white prints of the prime faces presented in random order. Participants rated only same-sex faces. To ensure that participants did not rate any familiar faces, participants in the rating study were selected from different schools than the participants of the priming experiments. Additionally, participants were explicitly instructed not to rate the faces of children they thought they might know. Subsequently, a mean attractiveness rating was computed for each face. The mean attractiveness ratings for faces serving as disliked, unknown, and liked primes were subjected to a one-way independent samples analysis of variance (ANOVA). The main effect of prime type was not statistically significant, $F = 0.16$ (mean attractiveness rating across categories = 2.33), demonstrating that there were essentially no between-category differences in the facial attractiveness of the prime stimuli.

Experimental sessions and design. Participants were tested individually in a quiet room at their school with a transportable stimulus presentation and response-gathering system by the first author and a trained undergraduate research assistant. Before a testing session, participants were told that they would be presented a sequence of two pictures of faces on each trial. They were instructed to fixate at the center of the screen throughout the trials, ignore the first picture (the prime), and to concentrate on categorizing, as fast as possible, the facial expression of the second picture (the probe) as angry or happy. Before the experiment, participants were shown examples of angry and happy facial expressions and presented with 10 practice trials.

Figure 1a shows the sequence of events on each trial. A trial started with a presentation of the central fixation cross for 500 ms. Next, the prime picture (a disliked, neutral, or liked peer) was displayed for 50 ms, followed by a 50-ms display of a structural grayscale mask (low visibility) or a blank screen (high visibility), depending on the type of the trial. Finally, the picture depicting angry or happy facial expression was displayed until the participant pressed the response button. The response buttons (i.e., left vs. right button for angry vs. happy expressions) were counterbalanced across participants. Each participant performed two blocks of the task, each consisting of 28 trials of each type in random order totaling 168 trials per block and a grand total of 336 trials. A short break was held between the two blocks. After the experiment, participants were asked to nominate up to three peers they liked and disliked the most in order to control for the stability of the relationship schemas (i.e., liking vs. disliking) and asked whether they knew the adolescents serving as neutral (unknown) primes. The experimental conditions were thus combined into 2 (visibility: low vs. high) × 3 (prime–probe congruency: incongruent vs. neutral vs. congruent) fully within-subjects design.

Results

To verify the stability of the relationship schemas, we assessed the congruency of peer nominations collected 1 month before and immediately after the experiment by examining whether children nominated the same targets among their liked and disliked peers. On average, the stability was high (79%). We

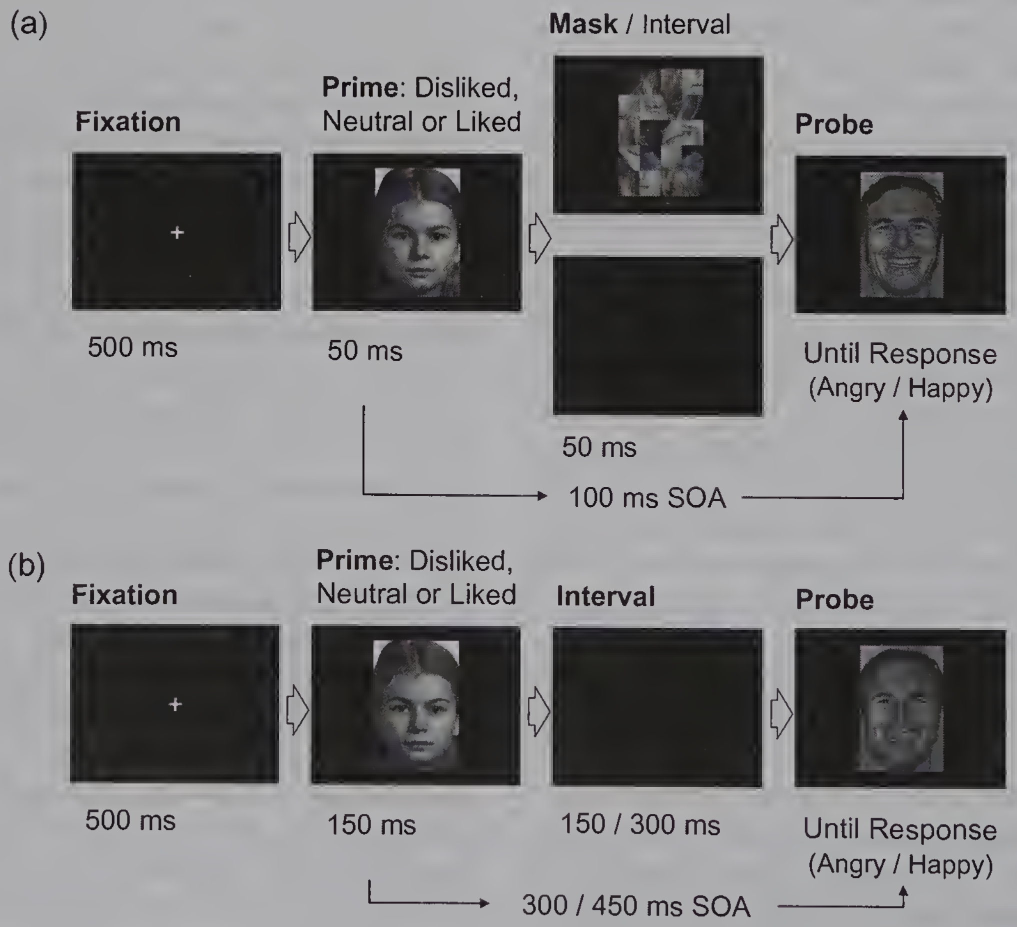

(a)
Fixation
Prime: Disliked, Neutral or Liked
Mask / Interval
Probe
500 ms
50 ms
50 ms
Until Response (Angry / Happy)
100 ms SOA
(b)
Fixation
Prime: Disliked, Neutral or Liked
Interval
Probe
500 ms
150 ms
150 / 300 ms
Until Response (Angry / Happy)
300 / 450 ms SOA

Figure 1. Sequence of events on a trial in Experiments 1a and 2b.
Note. SOA = stimulus onset asynchrony. Facial expression probes from Ekman and Friesen (1976). Copyright 1976 by P. Ekman and W. V. Friesen.

also compared whether nominations for liking versus disliking were equally stable by repeated measures t tests and found that the stability was slightly higher for liking (88%) than for disliking (70%) peers, $t(23) = 2.10, p < .05$. On average, participants made 6% errors (i.e., categorized the probe picture wrongly) in the expression categorization task. RT data from correct responses were collapsed across the blocks. Anticipations (RTs < 80 ms), and retardations (RTs 2.5 *SD* above individual mean) were excluded from the data analysis according to the suggestions by Ratcliff (1993). These accounted for 4% of the trials. Next, participant-wise mean RTs were calculated for each Visibility × Prime–Probe Congruency condition. These data with pooled 95% confidence intervals (see Masson & Loftus, 2003) are presented in Figure 2. The mean RTs were subjected to a 2 (visibility: low vs. high) × 3 (prime–probe congruency: incongruent vs. neutral vs. congruent) repeated measures ANOVA. The results demonstrated a main effect of visibility, $F(1, 24) = 7.30, p < .05, \eta_p^2 = .23$, and prime–probe congruency, $F(1, 24) = 3.24, p < .05, \eta_p^2 = .12$. Moreover, the Visibility × Prime–Probe Congruency interaction proved significant, $F(2, 24) = 3.13, p < .05, \eta_p^2 = .12$. RTs were faster in the low visibility condition (834 vs. 840 ms, respectively). Additionally, planned comparisons (corrected for multiple comparisons) revealed that for the high-visibility condition, both incongruent, $F(1, 24) = 6.21, p < .05, \eta_p^2 = .21$, and congruent, $F(1, 24) = 9.00, p < .05, \eta_p^2 = .27$, probes were categorized more slowly than neutral probes. However, none of the comparisons reached significance ($F < 1$) in the low-visibility condition. Also, to control for potential gender moderation effects, we repeated the ANOVAs in this and all subsequent experiments with gender as a between-groups factor and verified that gender did not interact with the priming effects in any of the experiments.

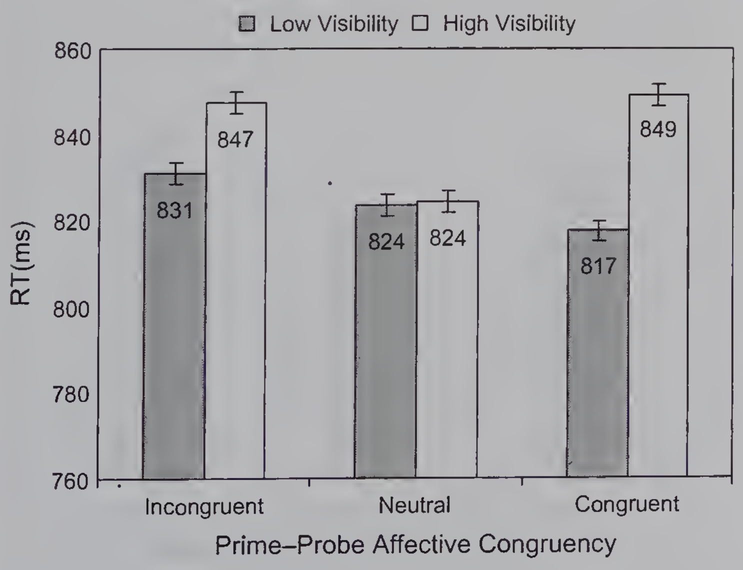


Figure 2. Mean response latencies (ms) with pooled 95% confidence intervals as a function of prime visibility and prime–probe affective congruency in Experiment 1.

Discussion

In line with previous research using affective pictorial primes and probes (Hermans, Spruyt, De Houwer, & Eelen, 2003, Experiment 1), no affective priming effect was found for the low-visibility condition. In contrast, the data for the high-visibility condition suggest that the relational schemas might have been activated by the primes. However, the results did not show the traditional affective priming effect. Instead, longer RTs were demonstrated after congruent as well as incongruent prime–probe pairs (i.e., after emotional primes) as compared to the probes preceded by neutral primes. This somewhat surprising finding becomes understandable if we consider the processing stream of face perception in more detail. The widely accepted models of face recognition (Bruce & Young, 1986; Haxby et al., 2000) assume that structural encoding of faces, which is necessary for identity recognition is undertaken prior to the retrieval of knowledge related to that person. The emotional reactions potentially arising from the stimulus faces in the current experiment were contingent on the recognition of the *identity* of the face: Across the different prime categories, the stimuli had equally neutral expressions and were rated as equally attractive; thus, they could be considered as *intrinsically neutral* stimuli. However, each participant saw faces of peers that they considered as pleasant and unpleasant due to their relationship history with the persons depicted in the stimuli. Consequently, it is likely that both the facial identity recognition and the *subsequent* retrieval of the affective information (i.e., relational schemas) attached to the perceived individual could not be undertaken during the short 100-ms period between the prime and the probe stimulus demanding response. This argument is supported by the fact that facial identity priming is usually demonstrated at much longer SOAs (up to 700 ms; see Bruce & Valentine, 1986) than affective priming (around 300 ms; see Hermans et al., 2001). In the current experiment, it is likely the participants were able to *detect* the liked and disliked primes as familiar (i.e., either pleasant or unpleasant) but could not undertake the more elaborate *discrimination* of the stimulus identity as liked or disliked. Subsequently, both liked and disliked primes resulted in interference and led to slower responses in both congruent and incongruent priming conditions as compared to the neutral priming condition. This is also in line with the model of postattentional affective processing by Storbeck et al. (2006), which suggests that the features of any object must first be integrated, and the object itself identified prior to affective analysis. In other words, postattentional affective priming would first involve the facial identity recognition, which would subsequently be followed by the affective analysis of the target face.

Experiment 2

If the affective priming effect is contingent on the recognition of the facial identity, it could be expected that affective priming would be manifested if the duration of the prime display was longer. We conducted Experiment 2 to test this hypothesis. We increased the prime display duration to 150 ms and replaced the stimulus visibility manipulation with a SOA (300 vs. 450 ms) manipulation. The selection of the SOAs was based on the SOAs used in the facial identity priming paradigms (Bruce & Valentine, 1986) as well as recent findings with the affective priming paradigm that have demonstrated that, though affective priming is automatic, affective processing is postattentive (Calvo & Nummenmaa, 2007). If the priming effect would only be contingent on the resources available for the visual processing of the prime, we expected that priming effects would be manifested at both SOAs. However, if the priming would be contingent on facial identity recognition and subsequent activation of affective relational schemas, priming effects were expected to emerge only at the 450-ms SOA.

Method

Participants

Thirty 6th-grade adolescents (10 girls, mean age = 13 years) not participating in Experiment 1 served as participants subjects and were compensated with a movie ticket. All the participants were randomly chosen from the same participant pool as those in Experiment 1. Of the 31 adolescents originally participating in the study, one had to be removed from the data analysis due to a chance-level performance (56% correct) in the expression recognition task. With regard to measures of rejection, withdrawal, and aggression, there were no differences between the children who participated in Experiment 2 and those who did not participate in any of the three experiments ($N = 180$, $t \leq 1.26$).

Apparatus, Stimuli, Experimental Design, and Procedure

Apparatus, stimuli, experimental design, and procedure were similar to those in Experiment 1, with three exceptions. First, the prime display duration was increased to 150 ms. Second, instead of the structural mask, a blank screen was presented after the primes on all trials. Third, two different prime–probe SOAs (300 vs. 450 ms) were used (see Figure 1b for events in trials). This resulted in a 2 (SOA: 300 vs. 450 ms) × 3 (prime–probe congruency: incongruent vs. neutral vs. congruent) fully within-subjects design.

Results

Again, the average stability of the affect felt toward the peers was high (73%), and it was slightly higher for liked (81%) than for disliked (65%) peers, $t(29) = 2.40$, $p < .05$. There were no differences in the facial attractiveness of the disliked, unknown, and liked prime stimuli, $F = 1.40$. The mean attractiveness rating across categories was 2.32. Mean RTs (an average of 7% errors and 2.5% of anticipations and retardations omitted) for the different SOA × Prime–Probe congruencies are presented in Figure 3. The mean RTs were analyzed with a 2 (SOA: 300 vs. 450 ms) × 3 (prime–probe congruency: incongruent vs. neutral vs. congruent) repeated measures ANOVA. The main effect for SOA reached significance, $F(1,29) = 40.55$, $p < .05$, $\eta_p^2 = .58$, demonstrating the classic foreperiod effect (see Niemi & Näätänen, 1981). RTs were faster for 450-ms than for 300-ms SOAs (794 vs. 833 ms, respectively). Although the main effect of

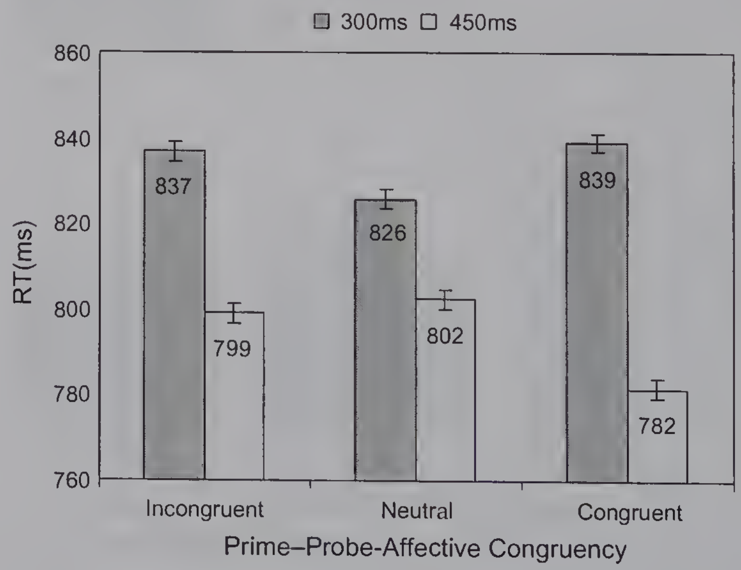


Figure 3. Mean response latencies (ms) with pooled 95% confidence intervals as a function of stimulus onset asynchrony and prime–probe affective congruency in Experiment 2.

prime – probe congruency did not reach significance, $F = 1.26$, the SOA × Prime – Probe Congruency interaction was significant, $F(2, 58) = 3.85, p < .05, \eta_p^2 = .12$. Multiple comparisons revealed that this interaction resulted from the fact that at the 450-ms SOA, congruent probes were categorized faster than neutral, $F(1, 29) = 6.57, p < .05, \eta_p^2 = .19$, and incongruent, $F(1, 29) = 4.75, p < .05, \eta_p^2 = .30$, probes (RTs = 781, 802, and 798 ms, respectively). However, neutral probes were not categorized faster than the incongruent probes, $F < 1$. There were no differences in RTs between congruent, neutral, and incongruent probes at the 300 ms SOA, $F < 1$.

As the congruency variable was pooled across liked and disliked primes, we also wanted to analyze whether the liked and disliked primes would produce equally strong priming effects. For that purpose, we compared the priming scores for happy (i.e., $\mathrm{RT}_{\text{disliked – happy}} - \mathrm{RT}_{\text{liked – happy}}$) and angry ($\mathrm{RT}_{\text{liked-angry}} - \mathrm{RT}_{\text{disliked – angry}}$) facial expressions with each other on the 450-ms SOA condition. Both priming scores were positive (22 vs. 12.50 ms) showing that disliked primes facilitated recognition of angry expressions and liked primes facilitated recognition of happy expressions, but there was no statistically significant difference between the priming scores for happy versus angry expressions, $t < 1$. This suggests that disliked and liked primes speeded up the categorization of congruent facial expressions to an equal extent.

Discussion

The main finding of Experiment 2 was that an affective priming effect was demonstrated when the prime – probe SOA was 450 ms. RTs were faster for congruent (disliked / angry or liked / happy) than for neutral or incongruent prime – probe pairs. This implies that the affective component of relational schemas is activated automatically—albeit not very rapidly—upon perception of a familiar peer. The results suggest that the activation of peer-relational affective information resulting from face perception is contingent on postattentional recognition of facial identity. The SOA manipulation revealed that the priming effect was crucially dependent on the total processing time and not the prime display time. Namely, the prime display time was similar (150 ms) for both SOAs, but the affective priming effect only emerged when the prime – probe SOA was 450 ms. Accordingly, presenting a face for 150 ms was definitely enough for identity recognition, but the affective priming effect took approximately 450 ms to occur. The time course of the observed priming effect was longer than what is typically observed in affective priming studies. The maximal affective priming effects are usually observed at 150 – 300 ms SOAs (see Hermans et al., 2001), although recently Carroll and Young (2005) demonstrated affective priming effects for facial expressions at SOAs of 250 and 750 ms. However, facial identity priming effects are typically manifested at 700-ms SOA (Bruce & Valentine, 1986). As the occurrence of the priming effects in the current studies would require recognition of the facial identity of the depicted person, it is understandable that the time course of the observed priming effect resembles that of facial identity priming.

Experiment 3

Experiments 1 and 2 relied on priming facial expression recognition with faces of participants' peers. Although Experiment 2 demonstrated that relationship-specific affective reactions can be automatically elicited by a brief presentation of liked and disliked peers' faces, it remains open whether activation of relational schemas would influence other aspects of online SIP, such as interpretation of ambiguous situational cues, feelings of anger, or likelihood of retaliation. As outlined in the Introduction, SIP is usually studied by using hypothetical vignettes and asking questions which tap different steps of SIP. It has, however, been questioned (Orobio de Castro, 2004) whether such measures capture the automaticity of SIP.

To test the automaticity assumption, Experiment 3 implemented a computerized hypothetical vignette task with a priming manipulation. According to Crick and Dodge (1994), the way children handle peer provocation and rebuff situations is especially critical for their social adjustment in the peer group. Participants were thus presented, on the computer screen, with short vignettes (two or three sentences) describing either provocation or rebuff by a hypothetical anonymous same-sex protagonist. The protagonist's behavior had always a negative consequence for the participant, but the intent of the protagonist was displayed as ambiguous. The vignettes were followed by questions measuring attributions of hostility, experience of anger, and willingness to retaliate. Based on two assumptions—(a) relational schemas are activated automatically and (b) relational schemas affect the way adolescents process social information about the particular peer—we expected that after perceiving the disliked primes, participants would attribute more hostility, experience more anger, and be more likely to retaliate against the protagonist in the vignette.

Method

Participants

Thirty 6th-grade adolescents (10 girls, mean age = 13 years) participated in the experiment. The participants were randomly chosen from the same participant pool used in Experiments 1 and 2. All the participating children had parental permission to take part in the experiments. Again, no significant difference was found between the participants and the rest of the sample who did not participate in any of the three experiments ($N = 180$) on peer-nominated rejection, withdrawal, and aggression scores ($t \leq 0.97$).

Apparatus and Stimuli

Apparatus and prime stimuli were similar to those used in Experiments 1 and 2. Vignettes were presented on the computer screen, with 1 vignette occupying the screen at a time. Altogether 24 vignettes were used describing social interactions in which the participant was provoked (12 vignettes; e.g., "Pretend that you are playing soccer with other boys/girls. You turn away from the other players for a while, and suddenly somebody kicks the ball to your back. It hurts very much") or rebuffed (12 vignettes, e.g., "Pretend that you are in your classroom and you hear two boys/girls talking about a bike trip taking place next weekend. You would also like to join them. You go over and ask them if you could go to the trip as well. One of the boys/girls says 'No!'") by a hypothetical same-sex peer (protagonist).

Each vignette was followed by three questions that appeared on the screen one at a time: (a) Did the protagonist intend to harm you? (*attribution of hostility*; b) Would you be angry at him/her? (*experience of anger*) and (c) Would you do something to get even? (*retaliation*). A white response scale line measuring 16° of visual angle was presented below each question and horizontally centered on the screen. Verbal labels for the extreme values (*not at all* vs. definitely) of the response scale were presented at the corresponding ends of the response scale. Otherwise, the scale range was arbitrary to the participants. In terms of response recording, the scale actually ranged from −250 (*not at all*) to +250 (*definitely*). A red vertical line served as a cursor for answering the questions. It measured 3°, was centered vertically on the response scale, and could be moved along the response scale with the mouse. All stimuli were presented against a black background.

Procedure and Design

To identify the peers used as primes, we followed the same procedure as in Experiments 1 and 2. Testing was conducted at schools similarly to the previous experiments. Participants were told that the study concerned decision making in various social situations. They were told that they were going to see a series of short texts describing social interactions with peers, followed by questions about their own reactions to the situations. They were also informed that they would see some pictures of faces during the experiment, but it was stressed that the pictures were completely irrelevant to the actual experiment. Participants were instructed to fixate at the center of the screen at the beginning of the trials, ignore the picture of the face, and to read the subsequent vignette carefully and respond to the questions using the mouse. Figure 4 shows the sequence of events for each trial. A trial started with a presentation of the central fixation cross for 500 ms. Next, the prime picture (disliked, neutral, or liked peer) was displayed for 150 ms, followed by a 300-ms display of a blank screen, resulting in a SOA of 450 ms. Next, a vignette was presented on the screen until the participant double-clicked the mouse button, indicating that he or she had finished reading the vignette. After each vignette, the questions regarding attribution of hostility, experience of anger, and likelihood of retaliation were presented in a random order. Each question was presented on a separate display, and participants responded using the mouse to move the line cursor along the response scale. Before the experiment, participants were presented with four practice trials, followed by a single experimental block consisting of four trials of each type (Prime Type × Vignette Type) in random order totaling 24 trials, with the same three questions asked after each trial. The experimental factors were combined to a 3 (prime type: disliked vs. neutral vs. liked) × 2 (vignette type: provocation vs. rebuff) × 3 (question type: attribution of hostility vs. experience of anger vs. retaliation) fully within-subjects design.

Results

The average stability of the affect felt toward the peers was high (75%), and it was slightly higher for liked (79%) than for disliked (71%) peers, $t(29) = 2.50$, $p < .05$. The mean attractiveness rating for prime faces was 2.44, and there were no between-category differences in the perceived attractiveness of the disliked, neutral, and liked primes, $F = 0.5$. Subject-wise mean response scores (see Table 1) were computed for each condition in the vignette task. The scores were subjected to a 3 (prime type: disliked vs. neutral vs. liked) × 2 (vignette type: provocation vs. rebuff) × 3 (question type: attribution of hostility vs.

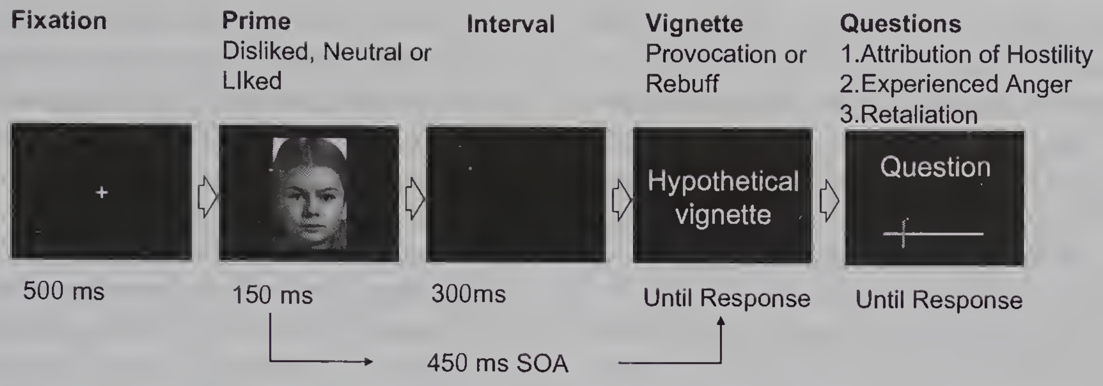


Figure 4. Sequence of events on a trial in Experiment 3.
Note. SOA = stimulus onset asynchrony.

experience of anger vs. retaliation) repeated measures ANOVA. This yielded significant main effects of prime type, $F(2, 58) = 3.38$, $p < .05$, $\eta_p^2 = .12$, and vignette type, $F(1, 29) = 4.02$, $p < .05$, $\eta_p^2 = .12$, as well as a Vignette Type × Question Type interaction, $F(2, 58) = 9.14$, $p < .05$, $\eta_p^2 = .24$. Multiple comparisons for the prime type showed that when compared to disliked, $F(1, 29) = 4.84$, $p < .05$, $\eta_p^2 = .14$, and neutral, $F(1, 29) = 4.54$, $p < .05$, $\eta_p^2 = .14$, primes, the liked primes resulted in less attributions of hostility, experience of anger, and retaliative responses. However, differences between disliked and neutral primes were nonsignificant, $F < 1$.

Planned comparisons for the Vignette Type × Question Type interaction demonstrated that for provocation vignettes, mean experience of anger was higher than attributions of hostility, $F(1, 29) = 10.74$, $p < .05$, $\eta_p^2 = .27$, and retaliation, $F(1, 29) = 37.99$, $p < .05$, $\eta_p^2 = .57$, and that mean attributions of hostility were higher than mean retaliation, $F(1, 29) = 4.84$, $p < .05$, $\eta_p^2 = .14$. For the rebuff vignettes, mean attributions of hostility were higher than experience of anger, $F(1, 29) = 5.23$, $p < .05$, $\eta_p^2 = .15$, and retaliation, $F(1, 29) = 44.84$, $p < .05$, $\eta_p^2 = .61$. Moreover, mean experience of anger was higher than mean retaliation, $F(1, 29) = 39.83$, $p < .05$, $\eta_p^2 = .58$. In addition, it was found that when compared to rebuff vignettes, provocation vignettes resulted in greater experience of anger, $F(1, 29) = 12.99$, $p < .05$, $\eta_p^2 = .31$, and retaliation, $F(1, 29) = 4.20$, $p < .05$, $\eta_p^2 = .13$, but no differences were found for attributions of hostility, $F < 1$.

Discussion

The results of Experiment 3 demonstrated that the automatic activation of relational schemas also had an effect on how adolescents processed social information other than facial expressions (cf. Experiment 2). In the case of disliked and neutral primes, participants attributed more hostility to the hypothetical peer, experienced more anger, and were more willing to retaliate as compared to liked primes. The results showed that nonretaliation was more common than retaliation across different priming conditions, suggesting that getting back at someone (even toward the disliked peer) might be infrequent. Still, we should be cautious when making such a claim as we did not have behavioral measures to assess retaliation (i.e.,

Table 1
Means and Standard Errors of Responses as a Function Prime Type, Vignette Type, and Question Type

	Disliked		Neutral		Liked	
	M	*SE*	*M*	*SE*	*M*	*SE*
Provocation vignettes						
Attributions of hostility	−4.43	17.13	−9.73	17.93	−11.90	15.87
Experience of anger	28.53	16.69	29.94	17.05	2.15	14.92
Retaliation	−62.19	17.73	−46.27	20.17	−71.62	15.65
Rebuff vignettes						
Attributions of hostility	−2.09	13.73	−3.51	14.71	−8.10	15.46
Experience of anger	−12.67	17.11	−22.66	16.16	−32.73	15.49
Retaliation	−71.08	16.73	−71.51	15.87	−95.87	13.32

Note. Ratings were provided on a scale that ranged from −250 (*not at all*) to +250 (*definitely*).

whether the participants actually retaliated against the peer) and associations between SIP steps and social behaviors are not necessarily very strong (for reviews, see Orobio de Castro et al., 2002; Yoon, Hughes, Gaur, & Thompson, 1999). Also, stronger retaliatory responses might have emerged if the sample had consisted of highly rejected or aggressive adolescents.

Overall, Experiment 3 corroborates our previous findings that relational schemas are automatically activated upon perception of a familiar peer. Additionally, as participants showed greater experience of anger for trials primed with disliked versus liked peers, the results provide additional support for our argument (Experiment 2) that perceiving a face of a familiar peer exerts an influence on subsequently arising affective reactions. Moreover, these findings also elaborate the results obtained in Experiment 2 in two ways. First, they show that the relationship-specific priming effects are not restricted to the affective domain but that activation of relationship schemas also influences interpretation of situational cues (intentionality) as well as response decision (likelihood of retaliation), which are two of the core components of the SIP model (Crick & Dodge, 1994). Second, the data suggest that the priming effects are relatively long lasting. Whereas a single trial in Experiments 1 and 2 lasted for less than 1 s, the average trial duration in Experiment 3 was 17.5 s. This underlines the importance of relationship schemas in guiding social behavior—although the schemas can be activated rapidly (in around 450 ms) they influence the social behavior at least for tens of seconds. It must, however, be noted that as the delay between the prime and the response was much longer in Experiment 3 than in Experiments 1 and 2, it is likely that participants had engaged in more reflective processing (see Orobio de Castro, 2004) in Experiment 3 when responding. Nevertheless, we argue that the priming effects have interacted with this reflective processing in automatic fashion, as the primes were presented only briefly and were irrelevant to the vignette task.

General Discussion

We conducted three experiments to study automatic relationship-specific SIP. In Experiments 1 and 2, we primed angry versus happy facial expression recognition with the facial identity of liked, neutral, and disliked peers to examine whether the affective component of peer-relational schemas would be automatically activated. In Experiment 3, we presented similar identity primes before hypothetical provocation and rebuff vignettes in order to investigate whether automatic activation of peer-relational schemas would influence SIP. The affective priming effect found in Experiment 2 showed that when the prime and probe stimuli were congruent in terms of their affective valence, the categorization of facial expressions was faster when the stimulus display duration and prime–probe SOA were long (150 and 450 ms, respectively). In case of the short display duration (50 ms), presentation of both liked and disliked peers as primes slowed the categorization of the subsequent facial expression when the primes were masked with a blank screen. When structural masking was applied, no priming effects were manifested. Together Experiments 1 and 2 demonstrate that relationship-specific affective reactions are activated by the mere perception of a familiar face, if sufficient time is given between the presentation of the facial identity and the demanded response. This suggests that children do hold different views of different peers and that different affective reactions are associated with different individuals.

In addition, Experiment 3 demonstrated that automatically activated relationship-specific schemas influenced social-cognitive evaluations of hypothetical peers. More specifically, when the disliked peer served as a prime, participants attributed more hostility to the protagonist in the vignette, experienced more anger, and were more likely to retaliate. These findings have important implications for the (a) automaticity, (b) relationship specificity, and (c) emotional aspects of SIP. We will next address these issues in more detail.

Automaticity of SIP

An important question is whether the peer-relational schemas were activated automatically in the current experiments. Traditionally, automatic processes are assumed to be (a) fast and efficient (b) unintentional, and (c) unconscious (for review, see Moors & De Houwer, 2006). At the short display duration (Experiment 1), disliked and liked but not neutral primes resulted in interference only when masked with a blank screen. At first, this might seem to contradict the first assumption of automatic processing. However, this finding probably reflects latency due to the hierarchy of face recognition processes, that is, the fact that identity recognition precedes affective evaluation as discussed above. For the priming effects to occur, facial identity encoding must be undertaken prior to retrieving the peer-related affective associations from the long-term memory. Although this process can take up to 450 ms to accomplish, it might still be undertaken in automatic fashion. This argument is corroborated by the fact that the data seem to fulfill other criteria for automatic processing. First, the priming effects were unintentional in the sense that they occurred even though the prime faces were task irrelevant, and participants were instructed to ignore the prime face and concentrate on categorizing the facial expression of the probes or reading the vignettes and responding the subsequent questions. Second, automatic versus conscious (or volitional) priming results in different patterns of RTs. Posner and Snyder (1975) have demonstrated that automatic aspects of priming will produce facilitation for related targets with no inhibition for unrelated targets, whereas conscious anticipatory effects will produce facilitation for related *and* inhibition for unrelated targets. As the 450-ms SOA condition of Experiment 2 demonstrated facilitation without inhibition ($\mathrm{RT}_{\mathrm{congruent}} < \mathrm{RT}_{\mathrm{neutral}} = \mathrm{RT}_{\mathrm{incongruent}}$), we consider the priming effect to result from the automatic activation of the peer-relational schemas.

So far, although SIP in children and adolescents has been assumed to be highly automatic (Crick & Dodge, 1994), and priming has been often mentioned as a necessary method to study online SIP (e.g., Burks, Laird, Dodge, Pettit, & Bates, 1999; Crick & Dodge, 1994), empirical research on the issue is almost nonexistent. Thus, considering that our existing knowledge about SIP originates mainly from the studies using reflective measures (e.g., questions presented after vignettes), the present study adds valuable information by providing direct evidence for the automatic retrieval of relationship-specific information.

Relationship Specificity of SIP

In line with other recent studies showing relationship specificity in SIP (Burgess et al., 2006; Peets et al., 2007; Peets et al., 2008), our results suggest that adolescents have different representations of peers that, once activated, affect subsequent information processing. Although SIP patterns were initially regarded as trait-like characteristics (Crick & Dodge, 1994), there is growing evidence that information processing is much less rigid and is influenced by the target peer's reputation and the relationship between the child and the target peer (Dodge, 1980; Hymel, 1986; Peets et al., 2007). In addition, relationship information has been described as being organized in a hierarchical way (e.g., Bretherton, 1990), with children having a general working model of peers, or general peer beliefs (Rabiner, Keane, & McKinnon-Lewis, 1993; Rudolph et al., 1995; Salmivalli et al., 2005), as well as more specific relational schemas, or knowledge structures containing information specific to different peers and relationships with them. When children first meet an unfamiliar peer, they tend to rely on their general peer beliefs developed from past experiences within family and among peers. However, these beliefs can change as a result of the interactions children have with those peers (Rabiner et al., 1993). Thus, although some children and adolescents tend to have a generalized view of others as hostile, it does not necessarily mean that they could not have a positive experience with some peers and develop a positive view of those peers.

Affective Components of Relational Schemas

The results of the present study show that the affective nature of the relationship (i.e., liking vs. disliking a peer) is an important aspect of peer-relational schemas and thus support the notion by Lemerise and Arsenio (2000) that emotional information associated with different peers influences SIP. Specifically, the data suggest that the emotions associated with certain disliked or liked peers are prone to be activated automatically in social situations and will promote affect-congruent cognitions and, subsequently, behaviors. Given that emotions occur rapidly and involuntarily, we argue that these affect-congruent modulations in social cognitions and behavior are crucial for maintaining already-established relationships. It should be pointed out that although the representation of a disliked peer might also capture emotions such as disgust, fear, and sadness, anger was chosen as a probe most congruent with the negative prime in the present study. It is indeed likely that adolescents dislike some peers because they are *afraid* of them rather than feel anger toward them. In addition, as reasons for dislike become more heterogeneous with age (see Hayes, Gershman, & Halteman, 1996), anger might not be the core emotion felt toward the disliked peer. Whether the same effect would be obtained by using probes with different facial expressions should be empirically tested. This would enable more fine-grained analysis of the organization of the affective peer-relational schemas. In addition, our results suggest that the sociometric nomination procedure used widely in the research on peer relations is a valid method to be used to differentiate between the peers whom children have positive and negative representations of. As aggression is one of the main reasons during middle childhood and early adolescence why someone is disliked, it could be argued that differences between liked and disliked peers might be just a function of their general behavioral reputation. However, studies have shown that even when the aggressiveness of the target peers has been taken into account, affective nature of the relationship has still an effect on social-cognitive processes (Peets et al., 2007).

Moreover, research using reflective measures has shown that preadolescents' evaluations of different peers are congruent with their affect felt toward those peers (Peets et al., 2008). For example, in the case of disliked peers, they make more hostile attributions, expect less relational and instrumental outcomes for aggression, and have higher self-efficacy beliefs for aggression, with the opposite being true for liked peers. Experiment 3 further demonstrated that affect-congruent social-cognitive evaluations are automatic. We should point out, however, that we did not find different priming effects for disliked versus unknown peers. This was not totally unexpected. Studies in adults have demonstrated that unfamiliar individuals with prototypical neutral facial expressions are evaluated as mildly unpleasant instead of neutral (Lee, Kang, Park, Kim, & An, 2007). Accordingly, such unpleasantness of the neutral faces of strangers probably explains why the disliked and neutral primes result in similar effects in Experiments 2 and 3. In addition, it has been shown that general beliefs about unfamiliar peers are not associated with beliefs about familiar peers (Rabiner et al., 1993), suggesting that actual experiences within the peer group can shape children's initial beliefs. Thus, future research could test whether negative attitudes toward unknown peers become more negative or positive over time as a result of children's interactions with these peers. Moreover, as our results imply that adolescents might, in general, have more hostile than prosocial beliefs of strangers, future priming studies manipulating the facial expression (happy vs. neutral vs. angry) of the liked, neutral, and disliked peer primes should be conducted to assess how situational factors can modulate the effect of these beliefs on SIP.

Online SIP is assumed to affect which behavioral response is enacted in a given situation (the "sixth step" in the SIP model). Consequently, the practical significance of the current results is that the automatic activation of relationship-specific information could automatically trigger certain behavioral patterns concordant with this information. Experiment 3 showed that although there were differences between disliked and liked primes with regard to retaliation, adolescents were more likely to be nonretaliative than retaliative toward the hypothetical peer across different priming conditions. It is in accordance with a study by Peets et al. (2008), who also demonstrated that although it was common for preadolescents to hold different social-cognitive evaluations of their liked and disliked peers, self-reported aggression even toward disliked peers was rather infrequent. Thus, whether social cognitions translate into aggressive behavior might be dependent on the characteristics of the individual and/or the interaction partner. For instance, it may be that only aggressive individuals act on their aggression-encouraging cognitions toward the peers who are physically weak and rejected by other peers. It is possible that if we had studied, for instance, extremely aggressive children, their responses would have been more retaliatory than nonretaliatory in nature.

Our results confirm that affect is a core component of relationships: Children are inclined to be negatively biased toward their enemies and positively biased toward their friends (Parker & Gamm, 2003; Peets et al., 2007). However, one should note that in our experiments, the relationship primes were chosen on the basis of *unilateral* nominations. Mutual friendships and antipathies are relationships that have much more personal meaning and relevance to children than relationships characterized by unilateral positive or negative affect. It is thus likely that the magnitude of the priming effects would have been larger if we had selected targets characterized by mutual disliking or liking. Previous studies (e.g., Card & Hodges, 2007) have indeed shown that children are more likely to be recipients of aggression by mutual antipathies than by friends or peers characterized by unilateral dislike and that mutual dislike serves as a context in which negative experiences have the most detrimental effect on children's adjustment.

Although aggression is, to a great extent, relationship specific, there are individuals who have tendencies to be more aggressive across relationships (Coie et al., 1999; Peets et al., 2008). Future studies could examine whether automatic processing is different, for instance, for aggressive versus nonaggressive individuals. It is possible that due to their aversive experiences in the peer group, especially rejected-aggressive children and adolescents have hostility schemas, which are activated in the case of unknown peers. Over time, SIP might become less affected by situational cues and more influenced by the activation of already existing (relationship-specific) representations of peers. Burks et al. (1999, p. 803) have suggested that "the more ambiguous the environmental information, the more the individual has to rely on the chronically accessible construct, and, thus, the resulting representation is composed predominantly of information from the knowledge base." Consequently, automatic processes might serve as key factors explaining stability of the peer reputation and its resistance to change. For instance, when a child dislikes someone, he or she is prone to process the information about this peer so that it would be congruent with the negative affect felt toward the peer. When other members of the group (i.e., classmates) share this negative view, social-cognitive processes consistent with the representation might be even less likely to change. For instance, a study by Hymel (1986) demonstrated that negative behaviors by disliked peers were attributed to more stable causes than negative behaviors by liked peers. Thus, although the behavior of a rejected child could be modified, other peers might still interpret the new behavior in light of existing representations and respective social-cognitive processes. It could be expected that differentiated social-cognitive evaluations are even more stable when the classroom structure remains the same over time (which is often the case in Finnish schools).

Participants of our experiments were early adolescents. Although reasons for liking and disliking someone become more abstract and diverse when children get older (see Hayes et al., 1996), affect-congruent social cognitions are also present in older adolescents (see Hymel, 1986). Previous studies that had failed to find such effects among young children (e.g., McDowell et al., 2002) have utilized the traditional, reflective measures. Priming studies would provide an alternative way to examine whether relationship-specific schemas of peers start to emerge already early on but cannot yet be reflected upon.

Finally, our findings offer one possible explanation why relationship effects have been found to account for just as much of the variance in children's dyadic aggression as actor or target effects (Coie et al., 1999). Namely, our study suggests that the affective person-related information is also activated automatically upon perceiving a familiar individual and that the activated information guides both affective and SIP. Future research could examine whether there are also individual differences in terms of how automated this processing is. Although in the current study we used an affective priming technique, we would expect priming effects to be evident regardless of the specific priming method. Taken together, the current findings thus encourage further research on the relationship specificity and automaticity of social behaviors and on the mechanisms explaining it.

References

Adolphs, R. (2002). Recognizing emotion from facial expressions: Psychological and neurological mechanisms. *Behavioral & Cognitive Neuroscience Reviews*, *1*, 21–62.

Baldwin, M. W. (1992). Relational schemas and the processing of social information. *Psychological Bulletin*, *112*, 461–484.

Baldwin, M. W., Carrell, S. E., & Lopez, D. F. (1990). Priming relationship schemas: My advisor and the Pope are watching me from the back of my mind. *Journal of Experimental Social Psychology*, *26*, 435–454.

Banse, R. (1999). Automatic evaluation of self and significant others: Affective priming in close relationships. *Journal of Social & Personal Relationships*, *16*, 803–821.

Banse, R. (2001). Affective priming with liked and disliked persons: Prime visibility determines congruency and incongruency effects. *Cognition & Emotion*, *15*, 501–520.

Bargh, J. A. (1997). The automaticity of everyday life. In R. S. Wyer (Ed.), *The automaticity of everyday life: Advances in social cognition* (Vol. 10, pp. 1–61). Mahwah, NJ: Erlbaum.

Bodenhausen, G. V., Sheppard, L. A., & Kramer, G. P. (1994). Negative affect and social judgement: The differential impact of anger and sadness. *European Journal of Social Psychology*, *24*, 45–62.

Breitmeyer, B. G., & Ogmen, H. (2000). Recent models and findings in visual backward masking: A comparison, review, and update. *Perception & Psychophysics*, *62*, 1572–1595.

Bretherton, I. (1990). Communication patterns, internal working models, and the intergenerational transmission of attachment relationships. *Infant Mental Health Journal*, *11*, 237–252.

Bruce, V., & Valentine, T. (1986). Semantic priming of familiar faces. *Quarterly Journal of Experimental Psychology*, *38A*, 125–150.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, *77*, 305–327.

Burgess, K., Wojslawowicz, J., Rubin, K., Rose-Krasnor, L., & Booth-LaForce, C. (2006). Social information processing and coping strategies of shy/withdrawn and aggressive children: Does friendship matter? *Child Development*, *77*, 371–383.

Burks, V., Laird, R., Dodge, K., Pettit, G., & Bates, J. (1999). Knowledge structures, social information processing, and children's aggressive behavior. *Social Cognition*, *8*, 220–236.

Calvo, M. G., & Nummenmaa, L. (2007). Processing of unattended emotional visual scenes. *Journal of Experimental Psychology: General*, *136*, 347–369.

Card, N. A., & Hodges, E. V. E. (2007). Victimization within mutually antipathetic relationships. *Social Development*, *16*, 479–496.

Carroll, N. C., & Young, A. W. (2005). Priming of emotion recognition. *Quarterly Journal of Experimental Psychology*, *58A*, 1173–1197.

Coie, J. (1990). Toward a theory of peer rejection. In S. Asher & J. Coie (Eds.), *Peer rejection in childhood* (pp. 365–401). Cambridge, UK: Cambridge University Press.

Coie, J. D., Cillessen, A. H., Dodge, K. A., Hubbard, J. A., Schwartz, D., Lemerise, E. A., et al. (1999). It takes two to fight: A test of relational factors and a method for assessing aggressive dyads. *Developmental Psychology, 35,* 1179–1188.

Coie, J. D., Dodge, K. A., & Coppotelli, H. (1982). Dimensions and types of social status: A cross-age perspective. *Developmental Psychology, 18,* 557–570.

Crick, N., & Dodge, K. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological Review, 115,* 74–101.

Dion, K. K. (1973). Young children's stereotyping of facial attractiveness. *Developmental Psychology, 9,* 183–188.

Dodge, K. (1980). Social cognition and children's aggressive behavior. *Child Development, 51,* 162–170.

Dodge, K. (1986). A social information processing model of social competence in children. In M. Perlmutter (Ed.), *The Minnesota symposium on child psychology* (Vol. 18, pp. 77–125). Hillsdale, NJ: Erlbaum.

Dodge, K. A., Murphy, R. R., & Buchsbaum, K. (1984). The assessment of intention-cue detection skills in children: Implications for developmental psychopathology. *Child Development, 55,* 163–173.

Dodge, K., & Rabiner, D. (2004). Returning to roots: On social information processing and moral development. *Child Development, 75,* 1003–1008.

Dodge, K. A., & Somberg, D. R. (1987). Hostile attributional biases among aggressive boys are exacerbated under conditions of threats to the self. *Child Development, 58,* 213–224.

Dorsch, A., & Keane, S. P. (1994). Contextual factors in children's social information processing. *Developmental Psychology, 30,* 611–616.

Ekman, P., & Friesen, W. (1976). *Pictures of facial affect.* Palo Alto, CA: Consulting Psychologists Press.

Feldman, E., & Dodge, K. (1987). Social information processing and sociometric status: Sex, age, and situational effects. *Journal of Abnormal Child Psychology, 15,* 211–227.

Gobbini, M. I., Leibenluft, E., Santiago, N., & Haxby, J. V. (2004). Social and emotional attachment in the neural representation of faces. *NeuroImage, 22,* 1628–1635.

Harter, S., Waters, P., & Whitesell, N. R. (1998). Relational self-worth: Differences in perceived worth as a person across interpersonal contexts among adolescents. *Child Development, 69,* 756–766.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences, 4,* 223–233.

Hayes, D. S., Gershman, E. S., & Halteman, W. (1996). Enmity in males at four developmental levels: Cognitive bases for disliking peers. *Journal of Genetic Psychology, 157,* 153–160.

Hermans, D., De Houwer, J., & Eelen, P. (1994). The affective priming effect: Automatic activation of evaluative information in memory. *Cognition & Emotion, 8,* 515–533.

Hermans, D., De Houwer, J., & Eelen, P. (2001). A time course analysis of the affective priming effect. *Cognition & Emotion, 15,* 143–165.

Hermans, D., Spruyt, A., De Houwer, J., & Eelen, P. (2003). Affective priming with subliminally presented pictures. *Canadian Journal of Experimental Psychology, 57,* 97–114.

Hubbard, J., Dodge, K., Cillessen, A., Coie, J., & Schwartz, D. (2001). The dyadic nature of social information processing in boys' reactive and proactive aggression. *Journal of Personality and Social Psychology, 80,* 268–280.

Hymel, S. (1986). Interpretation of peer behavior: Affective bias in childhood and adolescence. *Child Development, 57,* 431–445.

Klauer, K. C., & Musch, J. (2003). Affective priming. Findings and theories. In J. Musch & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 7–49). Mahwah, NJ: Erlbaum.

Langlois, J. H., & Cookie, S. (1978). The effects of physical attractiveness and ethnicity on children's behavioral attributions and peer preferences. *Child Development, 48,* 1694–1698.

Larson, R., & Verma, S. (1999). How children and adolescents spend time across the world: Work, play, and developmental opportunities. *Psychological Bulletin, 125,* 701–736.

Lee, E, Kang, J. I., Park, I. H., Kim, J.-J., & An, S. K. (2007). Is a neutral face really evaluated as being emotionally neutral? *Psychiatry Research, 157,* 77–85.

Leibenluft, E., Gobbini, M. I., Harrison, T., & Haxby, J. V. (2004). Mothers' neural activation in response to pictures of their children and other children. *Biological Psychiatry, 56,* 225–232.

Lemerise, E., & Arsenio, W. (2000). An integrated model of emotion processes and cognition in social information processing. *Child Development, 71,* 107–118.

Lemerise, E. A., Gregory, D. S., & Fredstrom, B. K. (2005). The influence of provocateurs' emotion displays on the social information processing of children varying in social adjustment and age. *Journal of Experimental Child Psychology, 90,* 344–366.

Masson, M. E. J., & Loftus, G. R. (2003). Using confidence intervals for graphically based data interpretation. *Canadian Journal of Experimental Psychology, 57,* 203–220.

McDowell, D. J., Parke, R. D., & Spitzer, S. (2002). Parent and child cognitive representations of social situations and children's social competence. *Social Development, 11,* 469–486.

Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin, 132,* 297–326.

Niemi, P., & Näätänen, R. (1981). Foreperiod and simple reaction time. *Psychological Bulletin, 89,* 133–162.

Orobio de Castro, B. (2004). The development of social information processing and aggressive behavior: Current issues. *European Journal of Developmental Psychology, 1,* 87–102.

Orobio de Castro, B., Bosch, J. D., Veerman, J. W., & Koops, W. (2003). The influences of emotions regulation, attribution prompts, and delay, on response aggressiveness in antisocial boys. *Cognitive Therapy and Research, 27,* 153–166.

Orobio de Castro, B., Slot, N. W., Bosch, J. D., Koops, W., & Veerman, J. W. (2003). Negative feelings exacerbate hostile attributions of intent in highly aggressive boys. *Journal of Clinical Child and Adolescent Psychology, 32,* 57–66.

Orobio de Castro, B., Veerman, J. W., Koops, W., Bosch, J. D., & Monshouwer, H. (2002). Hostile attribution of intent and aggressive behavior: A meta-analysis. *Child Development, 73,* 916–934.

Parker, J. G., & Asher, S. R. (1987). Peer relations and later personal adjustment: Are low-accepted children at risk? *Psychological Bulletin, 102,* 357–389.

Parker, J. G., & Gamm, B. K. (2003). Describing the dark side of preadolescents peer experiences: Four questions (and data) on preadolescents' enemies. In E. V. E. Hodges & N. A. Card (Eds.), *Enemies and the darker side of peer relations* (pp. 55–72). San Francisco: Jossey-Bass.

Peets, K., Hodges, E. V. E., Kikas, E., & Salmivalli, C. (2007). Hostile attributions and behavioral strategies in children: Does relationship type matter? *Developmental Psychology, 43,* 889–900.

Peets, K., Hodges, E. V. E., & Salmivalli, C. (2008). Affect-congruent social-cognitive evaluations and behaviours. *Child Development, 79,* 170–185.

Posner, M., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola symposium* (pp. 55–85). Hillsdale, NJ: Erlbaum.

Rabiner, D., Keane, S., & McKinnon-Lewis, C. (1993). Children's beliefs about familiar and unfamiliar peers in relation to their sociometric status. *Developmental Psychology, 29,* 236–243.

Rabiner, D., Lenhart, L., & Lochman, J. (1990). Automatic versus reflective social problem solving in relation to children's sociometric status. *Developmental Psychology, 26,* 1010–1016.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin, 114,* 510–532.

Ray, G. E., & Cohen, R. (1997). Children's evaluations of provocation between peers. *Aggressive Behavior, 23,* 417–431.

Rudolph, K., Hammen, C., & Burge, D. (1995). Cognitive representations of self, family- and peers in school-age children: Links with social competence and sociometric status. *Child Development, 66,* 1385–1402.

Salmivalli, C., Ojanen, T., Haanpää, J., & Peets, K. (2005). "I'm OK but you're not" and other peer-relational schemas: Explaining individual differences in children's social goals. *Developmental Psychology, 41,* 363–375.

Storbeck, J., Robinson, M. D., & McCourt, M. E. (2006). Semantic processing precedes affect retrieval: The neurological case for cognitive primacy in visual processing. *Review of General Psychology, 10,* 41–55.

Yoon, J., Hughes, J., Gaur, A., & Thompson, B. (1999). Social cognition in aggressive children: A metaanalytic review. *Cognitive and Behavioral Practice, 6,* 320–331.

Child Development, November/December 2008, Volume 79, Number 6, Pages 1676–1692

Romantic Partner Selection and Socialization During Early Adolescence

Valerie A. Simon
Wayne State University

Julie Wargo Aikins
University of Connecticut

Mitchell J. Prinstein
University of North Carolina at Chapel Hill

This prospective study examined romantic partner selection and socialization among a sample of 78 young adolescents (6th–8th graders). Independent assessments of adolescent and romantic partner adjustment were collected before and after relationships initiated via peer nomination and self-report. Prior to their relationship, adolescents and partners were significantly alike on popularity, physical attraction, and depressive symptoms. Controlling for initial similarity, partners' popularity, depressive symptoms, relational aggression, and relational victimization significantly predicted changes in adolescents' functioning in these areas over time. However, the magnitude and direction of change varied according to adolescents' and partners' prerelationship functioning. In general, adolescents who dated high-functioning partners changed more than those who dated low-functioning partners, and partner characteristics predicted greater change among low- versus high-functioning adolescents. Results were consistent even when controlling for best friend characteristics. The current findings are among the first to demonstrate unique contributions of romantic partner characteristics to adolescents' psychosocial functioning.

The emergence of romantic relationships is among the most significant psychosocial developments of adolescence. Romantic experiences appear to be associated with a diverse range of psychosocial outcomes. Some studies document enhancements in social status, romantic self-concept, self-esteem, and feelings of belonging, whereas others link romantic involvement to higher levels of depression, mood swings, conflict, and antisocial behavior (for reviews, see Collins, 2003; Furman & Shaffer, 2003). The processes by which romantic relationships could affect adolescent adjustment are not well understood. Characteristics of romantic partners are believed to be an important factor but little is known about adolescents' choice of romantic partners or the consequences of romantic partner characteristics on psychosocial functioning (Collins, 2003). The current study addresses this important gap in the literature by examining partner selection and socialization in adolescents' romantic relationships.

At all ages, partner selection and socialization processes appear to promote homophily between relationship partners (McPherson, Smith-Lovin, & Cook, 2001). A general tendency to be attracted to similar others increases the odds of choosing partners who share commonalities. Once the relationship is formed, partners tend to shape and reinforce similarities over time. Among youth, homophily has been documented in friendships and peer groups (Espelage, Holt, & Henkel, 2003; Hogue & Steinberg, 1995). In adulthood, homophily has been reported among dating, cohabitating, and married couples (Blackwell & Lichter, 2004; McPherson et al., 2001).

To the extent that homophily pervades social relationships, we might also expect to find homophily within adolescents' romantic relationships. If true, adolescents and partners would be similar to one another prior to relationship formation (selection) and become more similar over time (socialization). Yet we know of no published studies on this topic. Findings of adult romantic homophily are suggestive; however, there are likely to be developmental differences in the partner characteristics that are salient to adolescent and adult relationships. Research on peer influence offers insights about which partner characteristics could be salient during adolescence, yet

---

This research was supported by grants from National Institute of Mental Health (R01-MH59766) and the American Foundation of Suicide Prevention awarded to M.J.P. We would like to thank Annie Fairlie, Robin M. Carter, Daryn David, Carrie Hommel, and Erica Foster for their assistance with data collection and to all the adolescents and families who participated in this project.

Correspondence concerning this article should be addressed to Valerie A. Simon, Department of Psychology, Merrill-Palmer Institute, Wayne State University, 71 East Ferry Street, Detroit, MI 48202. Electronic mail may be sent to vsimon@wayne.edu.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7906-0010

friend, peer group, and romantic relationships have distinctive features (e.g., sexuality) and are thus not inherently redundant. Studies of adolescents' romantic partner preferences are informative regarding the characteristics adolescents find appealing (e.g., Hansen, 1977; Regan & Joshi, 2003), but preferences do not necessarily predict adolescents' actual choices or behavior. Understanding how adolescents select romantic partners and the contributions of romantic partner characteristics to adolescents' psychosocial functioning would provide important information about individual differences in links between romantic involvement and adjustment. In the current study, we assessed prerelationship similarities among prospective romantic partners and examined whether romantic partners' prerelationship characteristics predicted adolescents' subsequent psychosocial functioning.

Romantic Partner Selection

Studies of interpersonal attraction and adult romantic homophily have consistently identified status dimensions as important to partner selection (McPherson et al., 2001; Regan & Joshi, 2003). Status dimensions include sociodemographic factors (e.g., age, ethnicity) as well as ascribed characteristics, such as social standing or physical attractiveness. The current study focused on two ascribed status variables that are salient to young adolescents—physical appearance and popularity. Adolescents rate appearance as important to romantic partner selection, and adult partners' self-ratings of attractiveness tend to be positively related (Feingold, 1988; Regan & Joshi, 2003; Roscoe, Diana, & Brooks, 1987). Accordingly, we predicted that adolescents would select romantic partners who are similar to themselves in physical attractiveness and body appeal. Just as friends tend to be similar in social status, we expected that adolescents would also select partners who shared similar levels of popularity (Haselager, Hartup, van Lieshout, & Riksen-Walraven, 1998).

The function and structure of young adolescents' peer groups could further promote similarity between romantic partners. Because friendships and romantic relationships serve comparable functions, adolescents could use similar criteria to select friends and romantic partners (Furman, Brown, & Feiring, 1999). In addition, early romantic relationships are initiated within mixed-sex peer groups (Connolly, Craig, Goldberg, & Pepler, 2004), which are themselves subject to homophily effects at both the friend and the peer-group level (Espelage et al., 2003; Hogue & Steinberg, 1995). Hence, similarity could be particularly evident during early adolescence, when adolescents and partners are more likely to both belong to peer groups that already share similar features. Friends could also encourage romantic partner similarity through their roles as judges and brokers of romantic relationships. By defining what makes for suitable partners, friends could actively discourage the selection of partners who are discrepant in developmentally salient social or behavioral norms, such as social status or aggression (Brown, 1999). Such counsel could be especially prescriptive for young adolescents who are nearing peak susceptibility to peer conformity (Berndt, 1979; Krosnick & Judd, 1982; Steinberg & Silverberg, 1986).

For these reasons, we expected that adolescents would be similar to their romantic partners on dimensions that are important to friendships and peer groups at this age, including popularity, physical appearance, depressive symptoms, peer aggression, and peer victimization (Espelage et al., 2003; Kandel, Davies, & Baydar, 1990; Mariano & Harton, 2005; Prinstein, 2007; Stevens & Prinstein, 2005). However, similarity at the time of partner selection could be greater for observed than for psychological characteristics. After all, overt similarities are more easily identified than internal psychological characteristics during the initial attraction phase of relationships (Kandel, 1978; Urberg, Degirmencioglu, & Tolson, 1998). Accordingly, we predicted that prospective partners would be more similar in their physical appearance, peer popularity, peer aggression, and peer victimization than in their levels of depressive symptoms.

Implicit in these hypotheses is the suggestion that friends and romantic partners are alike. If adolescents select friends who are similar to themselves and romantic partners serve similar functions as friends, we would expect adolescents to choose romantic partners who are similar to their friends. Accordingly, we predicted that friends' and romantic partners' levels of popularity, physical appearance, peer aggression, and peer victimization would be significantly correlated. The salience of physical appearance to romantic attraction led us to anticipate that adolescents would be more similar to prospective romantic partners than friends on this particular dimension.

Romantic Partner Socialization

Adolescents may select peers to whom they are similar, but over time, these peers may also influence adolescents' behavior. For instance, friend characteristics predict changes in adolescents' depressive symptoms, peer aggression, and peer victimization over time, even after controlling for similarity (Espelage et al., 2003; Hogue & Steinberg, 1995; Jaccard, Blanton, & Dodge, 2005). In the current study, we examine whether romantic partner characteristics are predictive of similar changes in adolescents' psychosocial functioning. Such direct evidence of romantic partner socialization is currently absent from the literature. However, research by Adams and colleagues indicates that young adolescents rate romantic partners to be at least as influential as friends on their thoughts, feelings, and behavior (Adams, Laursen, & Wilder, 2001). These findings, coupled with the time and emotion adolescents invest in romantic relationships, suggest that partners could be important socializing agents.

Researchers typically estimate peer socialization by assessing whether peers' functioning at one time significantly predicts youths' functioning at a later time. The current study employed a similar design. We assessed whether partners' psychosocial functioning prior to the relationship predicted change in adolescents' psychosocial functioning after the two began dating. Changes in adolescents' psychosocial functioning predicted by partners' prerelationship functioning are interpreted as evidence of partner socialization.

We expected that some partner characteristics would be important to selection or socialization, whereas others would be important to both. As noted earlier, appearance and popularity should each be important to partner selection. Yet only partners' popularity was expected to exert a significant socialization effect when predicting adolescents' functioning over time. To the extent that status grading is an important aspect of early romantic relationships, partners' prerelationship popularity should be positively associated with adolescents' subsequent popularity (Skipper & Nass, 1966). Prior findings of significant peer socialization effects on adolescents' depressive symptoms, peer aggression, and peer victimization suggested that romantic partners' functioning in these areas could also be important predictors. Furthermore, we expected that some partner characteristics would predict adolescents' subsequent functioning even in the absence of initial partner similarity. For instance, covert characteristics, such as internalizing symptoms, could be less important to partner selection but then gain influence as partners spend time together (Baker, Milich, & Manolis, 1996; Stevens & Prinstein, 2005).

In the majority of homophily studies, peer socialization is examined in an additive model where peer characteristics are tested as main effects under the assumption that all peers influence all adolescents in a roughly equivalent manner. More recently, Hartup (1999, 2005) has argued that peer socialization is best conceived as an interaction between characteristics of the socializing agent and the socialized individual. In other words, some partners could be more influential than others, and some adolescents could be more open to influence than others. Relatively few studies have examined peer influence from this perspective; however, those that have suggested that healthy and negative peer influences are contingent on both partners' initial level of functioning (e.g., Adams, Bukowski, & Bagwell, 2005; Dishion & Owen, 2002).

Similarly, we hypothesized that high- and low-functioning romantic partners might predict different patterns of change for high- and low-functioning adolescents. For example, dating a partner with few psychosocial problems (i.e., high functioning) could be especially helpful for adolescents experiencing more psychosocial problems (i.e., low functioning) but could be less useful for adolescents with few problems. In support of this idea, having a friend who is low on aggression predicts decreases in adolescents' aggression over time, but only for adolescents who are initially more aggressive (Adams et al., 2005). In the current study, we expected that adolescents with more prerelationship problems who dated high-functioning partners would show improvements in psychosocial functioning over time. Adolescents with few prerelationship problems who paired with similarly high-functioning partners were not expected to change.

In contrast, the characteristics of low-functioning partners (i.e., those with more prerelationship problems) were not expected to predict much change in adolescents' psychosocial functioning. High-functioning adolescents, who have few psychosocial problems, could be resistant to the problems of their low-functioning partners (Adams et al., 2005; Dishion & Dodge, 2005). On the other hand, low-functioning partners could reinforce the problems of low-functioning adolescents once the two begin dating (Duncan, Boisjoly, Kremer, Levy, & Eccles, 2005; Poulin, Dishion, & Burraston, 2001). In either case, we expected to see little change among adolescents dating low-functioning partners.

The Current Study

Although romantic partners are presumed to affect adolescents' development, we know little about how adolescents select romantic partners or the consequences of their choices. This could be partly due to methodological challenges, as isolating selection from socialization effects to predict change as a function of partner characteristics requires longitudinal designs that can identify individuals' future romantic partners. In the current study, we identified adolescent couples within a longitudinal school-based sample. Using a follow-back design, we traced adolescents and romantic partners to a prior data collection to examine prerelationship similarities. This allowed for a relatively pure estimation of selection effects, as neither adolescents nor partners identified themselves as being romantically involved at that time. Socialization effects of romantic partners were then estimated by using partners' prerelationship characteristics to predict changes in adolescents' psychosocial functioning over the time prior to (Time 1) and after relationships were established (Time 2).

Young adolescents' friend and romantic relationships are frequently developed at school, and our school-based design allowed us to examine emerging romantic relationships in their broader social context. Although the data points are necessarily anchored around the initiation of romantic relationships, estimates of friendship similarity prior to the initiation of romantic relationships provide important information about the role of friends in romantic partner selection and socialization. Comparing the characteristics of adolescents' extant friends to those of prospective romantic partners as well as comparing the degree of similarity within friend and romantic dyads would offer new data on the relative importance of general and relationship-specific peer selection criteria in early adolescence. The inclusion of best friends also permitted us to examine romantic partner socialization while accounting for co-occurring socialization by best friends. We expected to find similarity but not redundancy across the two relationships. Romantic relationships are a unique facet of adolescent social development (see Collins, 2003; Furman & Shaffer, 2003, for reviews). Adolescents themselves report different expectations for friend and romantic relationships (Connolly & Goldberg, 1999), and empirical associations between the two relationships are significant but modest (Furman, Simon, Shaffer, & Bouchey, 2002).

Method

Participants

Participants included 78 adolescents (48% female) who were in the sixth (32 %), seventh (35%), and eighth (33%) grades at the beginning of the study. The ethnic composition of the sample included 87% European American, 2% African American, 4% Asian American, 2% Latino American, and 6% from mixed-ethnic backgrounds. Participants were enrolled in public schooling within a town of fairly homogenous middle-class socioeconomic status (per capita income = $32,301). According to school records, 11% of children were eligible for free/reduced lunch. These participants were part of a larger study (*n* = 520) designed to examine developmental trajectories of depressive symptomatology. All sixth-through eighth-grade students were invited to participate in the first phase of data collection for the larger study (Time 1). Consent forms were returned for 92% of families (*n* = 784); of these, 80% of parents gave consent for their child's participation (*n* = 637, 74% of total population). Students who were absent on one of the testing days (*n* = 10), provided incomplete data (*n* = 15), or refused to assent (*n* = 4) were excluded from analyses, yielding a final sample of 598 participants at Time 1. A total of 520 (87%) of these participants were available for testing 11 months later at Time 2, when students were in Grades 7 through 9. Attrition was due to participants moving away from the area (*n* = 36), absenteeism (*n* = 7), incomplete data (*n* = 30), and refusal to participate (*n* = 5). No significant difference was found between adolescents who participated in two versus one time points.

The current sample included adolescents who indicated on the Networks of Relationship Inventory (Furman & Buhrmester, 1985) that they began a romantic relationship between Time 1 and Time 2. Specifically, participants were asked whether they had a boyfriend or girlfriend, the name of this person, and how long this person had been their boyfriend or girlfriend. On average, these romantic relationships lasted 13.63 weeks (*SD* = 19.10). Because romantic partner and best friend data were needed to examine selection and socialization effects, only those adolescents whose best friend and romantic partner were also participants in the study were included in our sample. To avoid dependency in the data, one member of any reciprocally nominated romantic dyad was randomly dropped from the sample. This resulted in a data set in which each adolescent served as only a target participant or as a romantic partner. Similarly, no adolescent appeared as both a friend and a romantic partner in the data set. We did not drop any target adolescent who was named as another target adolescent's best friend because friendships were not the focus of the study and doing so could have resulted in a biased sample of adolescents whose best friends were not dating.

The 78 target participants were compared to 62 participants who reported a romantic relationship at Time 2 but who did not meet other study criteria (i.e., the best friend or romantic partner did not also participate in the study). No significant group differences emerged for gender, grade, or any Time 1 primary study variables. Similar analyses were conducted to compare adolescents in the larger sample who were and were not dating. Compared to nondating adolescents, students in romantic relationships were seen by peers as more physically attractive, $t(518) = 6.93$, $p < .001$, $M = 0.55$, $SD = 1.29$ versus $M = -0.10$, $SD = 0.89$; more popular, $t(518) = 6.68, p < .001, M = 0.53, SD = 0.98$ versus $M = -0.10$, $SD = 0.95$; less sad, $t(518) = -4.78, p < .001, M = -0.27$, $SD = 0.46$ versus $M = 0.06$, $SD = 1.08$; and less physically victimized by peers, $t(518) = -2.36$, $p < .05$, $M = -0.19$, $SD = 0.77$ versus $M = 0.04$, $SD = 1.06$.

Procedure

A letter introducing the study was mailed to the homes of all potential participants, and a consent form was sent home with each student. Parents were asked to either grant or deny their consent for their child's participation, and adolescents were asked to return the consent form regardless of their parent's decision. At both time points, questionnaires were administered to adolescents in their classrooms over 2 days. Each participant received a small token of appreciation (e.g., a key chain), a $5 gift certificate, and a raffle entry for a Sony Playstation 2 or Microsoft Xbox.

Measures

Popularity. Sociometric peer nominations were used to obtain measures of adolescents' peer perceived popularity at both time points. Students at this school were organized into teams about twice the size of traditional academic classrooms. Using an alphabetized roster of all classmates within their academic team, participants were asked to select an unlimited number of peers who were "most popular." The order of names was counterbalanced on these rosters (e.g., Z through A) to control for possible effects of alphabetization on nominee selection. A continuous measure of perceived popularity was computed by standardizing the number of most popular votes with higher scores indicating greater popularity (Rose, Swenson, & Waller, 2004).

Physical appearance. The peer nomination procedure described earlier also was used at both time points to obtain peer ratings of physical attractiveness and body appeal. The question for physical attractiveness was "Who is good looking?" and the question for body appeal was "Who is physically fit?" As described earlier, the number of nominations that each participant received for each question was summed and standardized relative to their teammates.

Peer-rated sadness. Peer nomination procedures were used to assess participants' sadness at each time point from the question "Who looks sad and unhappy most of the time?" The number of nominations received by each participant was summed and standardized relative to their teammates.

Self-rated depressive symptoms. The Children's Depression Inventory (CDI) was used to obtain self-reports of depressive symptoms at both time points (Kovacs, 1992). The CDI contains 27 items that assesses affective, cognitive, motivational, and somatic symptoms of depression. For each item, youth choose one of the three statements, scored 0 through 2, that best describe their level of depressive symptoms over the prior 2 weeks. Mean scores were computed for each student, with higher scores reflecting greater levels of depressive symptoms. In the current sample, Cronbach's alphas were .87 at Time 1 and .86 at Time 2.

Peer-rated aggression. Peer nomination procedures were used to assess participants' relational and physical aggression from the questions "Who ignores others or spreads rumors about others when they are mad at them?" (relational) and "Who starts fights, says mean things, and picks on others?" (physical). The number of nominations received by each participant for each item was summed and standardized relative to their teammates.

Peer-rated victimization. Peer nomination procedures were used to assess participants' relational and physical victimization from the questions "Who has lies, rumors, or mean things said about them?" (relational) and "Who gets picked on or teased by other kids?" (physical). The number of nominations received by each participant for each item was summed and standardized relative to their teammates.

Results

Preliminary Analyses

Ratings of depressive symptoms were obtained from adolescents as well as peers, and the two were significantly related ($r = .43$, $p < .001$). Prior to data analysis, scores for self-reported depressive symptoms were standardized to be consistent with sociometric measures and allow for meaningful comparisons across domains of functioning. Means and standard deviations for all measures are presented in Table 1 for target adolescents, romantic partners, and best friends. No significant mean differences were found between target adolescents', romantic partners', and best friends' corresponding values across any of these domains.

Partner Selection

Our first goal was to estimate romantic partner selection effects by examining prerelationship similarities in developmentally salient domains. Intraclass correlations between adolescents' and romantic partners' scores on each Time 1 variable indicated significant prerelationship similarities in four domains: popularity, body appeal, physical attractiveness, and self-rated depressive symptoms (see Table 2). To test the hypothesis that similarity would be greater for more observable characteristics, Fisher's r-to-z tests were calculated to compare the magnitude of similarity correlations across domains (Cohen, 1988). Among the four significant correlations, adolescent to romantic partner similarities were significantly larger for popularity than for physical attractiveness ($p < .001$) and self-rated depressive symptoms ($p < .05$). In addition, the similarity correlation for each of the four significant domains was significantly larger than each similarity correlation where no significant correlation was found (i.e., peer-rated sadness, physical aggression, relational aggression, physical victimization, and relational victimization; all $ps < .01$).

Parallels between adolescents' friend and romantic relationships had led us to hypothesize that these two relationship partners would share similar characteristics. The intraclass correlations between adolescents and their best friends on Time 1 variables paralleled those reported above for adolescents and prospective romantic partners, with the notable exception of body appeal (see Table 2). Specifically, adolescents were similar to both their best friend and romantic partner on popularity, attractiveness, and self-rated depressive symptoms. In addition, adolescents and best friends were significantly alike on Time 1 relational aggression. Comparing the magnitude of the similarity correlations across relationships revealed that adolescents were significantly more like friends than romantic partners on popularity, Williams's $t(75) = -2.65$, $p < .01$, and significantly more like romantic partners than friends on body appeal, Williams's $t(75) = 2.85$, $p < .01$. No significant differences in adolescents' similarity to romantic partners versus best friends were found for any of the other Time 1 variables. To further explore links between adolescents' friends and romantic partners, we computed Pearson correlations between the Time 1 characteristics of the two relationship partners (see Table 2). Adolescents' extant best friends and prospective romantic partners were significantly similar in their popularity, body appeal, attractiveness, and self-rated depressive symptoms at Time 1.

Partner Socialization

The second goal of this study was to estimate romantic partner socialization effects by testing whether partners' prerelationship characteristics predicted change in adolescents' functioning over time. We predicted that romantic partner characteristics would significantly predict adolescents' functioning at Time 2, even after accounting for co-occurring

Table 1
Means and Standard Deviations for All Primary Time 1 Variables

	Target adolescents		Best friends		Romantic partners	
Time 1 variables	Standardized	Unstandardized	Standardized	Unstandardized	Standardized	Unstandardized
Popularity	.38 (1.13)		.29 (1.19)		.58 (1.30)	
Body appeal	.37 (1.01)		.77 (2.95)		.60 (1.10)	
Physical attractiveness	.49 (1.23)		.51 (1.40)		.82 (1.42)	
Peer-rated sadness	−.21 (0.47)		−.23 (0.49)		−.28 (0.57)	
Self-rated depressive symptoms	.04 (1.01)	29 (0.29)	−.02 (1.02)	.22 (0.23)	.07 (1.02)	.28 (0.23)
Relational aggression	.09 (1.05)		.04 (0.98)		.19 (1.12)	
Physical aggression	.05 (1.13)		−.18 (0.56)		−.02 (0.78)	
Relational victimization	−.08 (0.80)		−.22 (0.65)		−.07 (0.67)	
Physical victimization	−.15 (0.74)		−.25 (0.48)		−.30 (0.33)	

Table 2
Correlations Among Time 1 Variables for Adolescents, Best Friends, and Romantic Partners

	Adolescent – romantic partner[a]	Adolescent – best friend[a]	Romantic partner – best friend[b]
Popularity	.56***	.68***	.46***
Body appeal	.42***	.07	.31***
Physical attractiveness	.25**	.23**	.24**
Peer-rated sadness	.01	.06	.11
Self-rated depressive symptoms	.38***	.35***	.19*
Relational aggression	.03	.27	.14
Physical aggression	−.14	.07	.15
Relational victimization	−.06	.07	−.07
Physical victimization	.09	.04	.09

[a] Intraclass correlation. [b] Pearson correlation.
$*p < .05$. $**p < .01$. $***p < .001$.

socialization by best friends. However, we expected that the strength and direction of the effects would vary according to both partners' and adolescents' prerelationship functioning.

Data analytic strategy. To test these hypotheses, we conducted a series of regression analyses to test the main and interactive effects of adolescents' and romantic partners' prerelationship characteristics on adolescents' psychosocial functioning at Time 2 (Baron & Kenny, 1986; Holmbeck, 2002). Prior to the analyses, all predictors were centered to reduce multicollinearity (Aiken & West, 1991). For each Time 2 outcome, adolescents' corresponding Time 1 scores were entered in Step 1 to assess stability over time. Best friends' Time 1 scores were entered in the second step to control for co-occurring socialization by friends. Romantic partners' Time 1 scores were entered in a third step to assess the main effect of romantic partner characteristics. The product term of adolescents' and romantic partners' Time 1 scores was entered in the final step to assess whether adolescents' and romantic partner's Time 1 functioning interacted to predict adolescents' Time 2 functioning.

All significant Adolescent × Partner interactions were subsequently probed following procedures described by Holmbeck (2002) in which slope estimates are calculated and examined at high ($M + 1\ SD$) and low ($M - 1\ SD$) levels of the moderator (Aiken & West, 1991). Significant interactions were probed twice, once with romantic partners' prerelationship functioning as the moderator and a second time with adolescents' prerelationship functioning as the moderator. This strategy allowed us to determine how change in adolescents' functioning over time varied according to partners' prerelationship functioning and by adolescents' prerelationship functioning. The two are related, though not identical, ways of decomposing and understanding the observed moderator effects.

Examining the moderating effects of romantic partner characteristics identifies whether change in adolescents' psychosocial functioning is greater when partners are high versus low functioning. Specifically, post hoc analyses assess the association between adolescents' functioning at Time 1 and Time 2 (i.e., stability) when partners are high and low functioning. Thus, change (i.e., lack of stability) in adolescents' characteristics over time as a function of romantic partners' characteristics is indicated by a nonsignificant slope (i.e., unstandardized beta). Examining the moderating effects of adolescents' prerelationship characteristics identifies whether romantic partner characteristics predict different amounts or types of change for high- versus low-functioning adolescents. Specifically, the post hoc analyses assess the association between partners' functioning at Time 1 and adolescents' functioning at Time 2 when adolescents were initially high versus low functioning. For these analyses, significant effects as a function of adolescents' initial functioning are indicated by a significant slope.

Controlling for co-occurring socialization by adolescents' best friends in Step 2 of the analyses provided a conservative test of romantic partner socialization. To the extent that friend and partner effects are confounded, this strategy could underestimate romantic partner socialization. To examine this possibility, we reran the regressions described above without controlling for best friend characteristics (i.e., eliminating Step 2 in the regressions). The pattern of results was very similar to that obtained when controlling for best friend characteristics, with almost identical beta weights and ΔR^2 values. No additional significant direct effects emerged for romantic partners' Time 1 characteristics. The similarity of results across the two sets of analyses suggested that romantic partner predictors were not redundant with friend predictors. Therefore, the results presented below are those from the analyses that include best friend values as a covariate in Step 2.

Popularity. Romanic partners' Time 1 popularity was not directly associated with adolescents' Time 2 popularity. However, as hypothesized, the interaction between adolescents' and partners' Time 1 popularity was a significant predictor (see Table 3). Results from post hoc probing of the moderating effects of romantic partners' popularity are plotted in Figure 1, where the lines depict the association between adolescents' popularity at Time 1 and Time 2 when partners' Time 1 popularity was high versus low. The beta coefficients were significant whether adolescents dated either high- or low-popular partners ($b = .75$, $p < .001$ and $b = 1.08$, $p < .001$, respectively), suggesting little change in adolescents' popularity over time. However, the beta coefficient was smaller for adolescents who dated high-popular partners, suggesting that these adolescents experienced more change in popularity than those who dated low-popular partners.

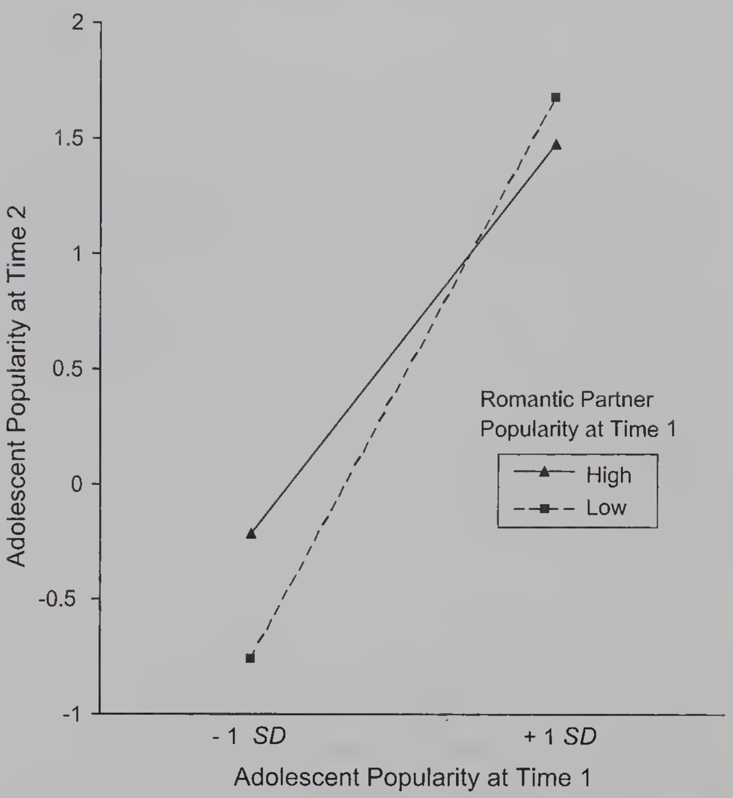


Figure 1. Graph of simple slopes for interactions between adolescent and romantic partner Time 1 popularity in predicting adolescents' Time 2 popularity with romantic partners' popularity as the moderator.

Results from post hoc probing of the moderating effects of adolescents' prerelationship popularity are plotted in Figure 2, where the lines depict the association between partners' popularity at Time 1 and adolescents' popularity at Time 2 when adolescents' Time 1 popularity was high versus low. Here, it can be seen that romantic partners' initial popularity predicted the subsequent popularity of low- but not high-popular adolescents ($b = .15$, $p < .05$ and $b = -.05$, $p = ns$, respectively). Among adolescents who were initially low in popularity, those who dated a high-popular partner were more popular at Time 2 than those who dated a low-popular partner.

Considered together, these two sets of post hoc analyses facilitate a better understanding of how the combination of adolescents' and partners' initial popularity predict change in adolescents' popularity over time. Figure 3 illustrates these patterns of change for low- and high-popular adolescents when they dated

Table 3

Summary of Hierarchical Regression Analyses Examining Main and Moderating Effects of Romantic Partner Characteristics on Adolescents' Popularity, Body Appeal, and Physical Attractiveness at Time 2

	Popularity		Body appeal		Physical attractiveness	
	β	ΔR^2	β	ΔR^2	β	ΔR^2
Step 1						
T1 adolescent (A) functioning	.87***	.76***	.91***	.80***	.71***	.65***
Step 2						
T1 best friend functioning	.14	.01	.07	.01	.20	.03
Step 3						
T1 romantic partner (RP) functioning	.08	.00	.12	.01	−.02	.00
Step 4						
T1 A × RP functioning	−.22	.02**	.05	.00	−.07	.00
Total R^2		.79***		.82***		.68***

Note. At each step, ΔR^2 is for the step and β is for the final model.
$^{**}p < .01$. $^{***}p < .001$.

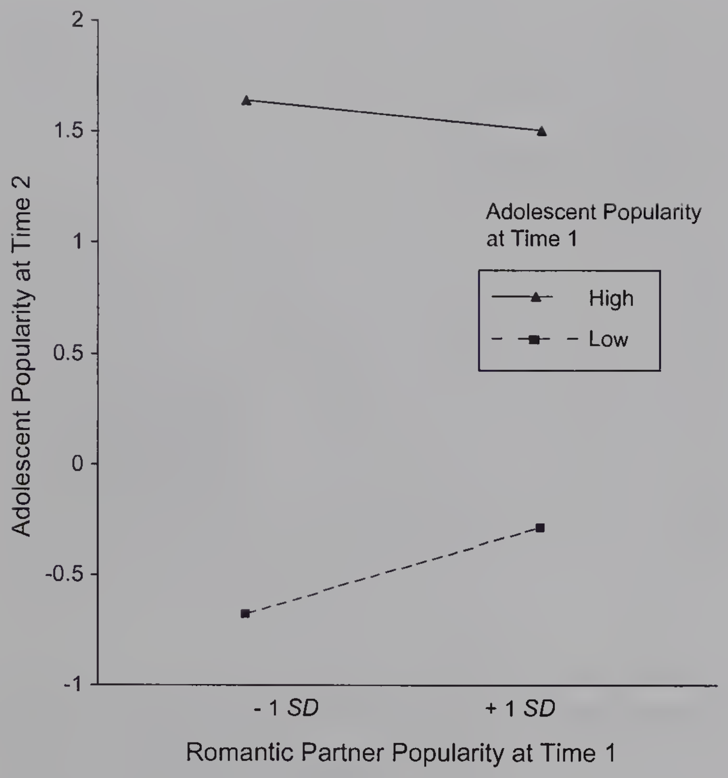


Figure 2. Graph of simple slopes for interactions between adolescent and romantic partner Time 1 popularity in predicting adolescents' Time 2 popularity with adolescents' popularity as the moderator.

low- versus high-popular partners. High and low functioning are defined in relation to the sample mean ($M \pm 1$ SD). Adolescents who were highly popular at Time 1 remained fairly stable over time, regardless of their partners' popularity. In contrast, low-popular adolescents changed, but this was limited to those who dated high-popular partners. Low-popular adolescents who dated a high-popular partner became more popular, whereas low-popular adolescents who dated a low-popular partner did not change.

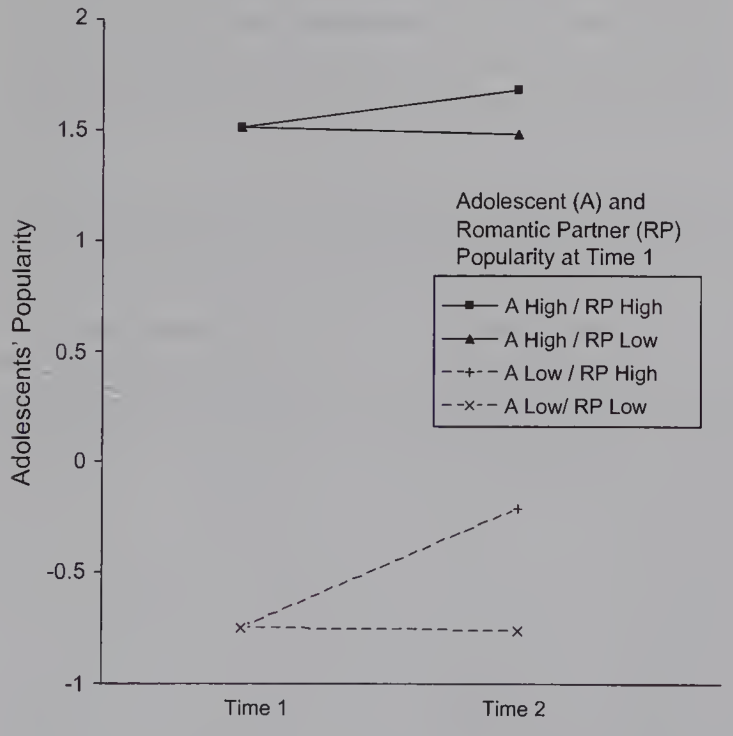


Figure 3. Graph of change patterns in adolescents' popularity from Time 1 to Time 2 by combination of adolescent and partner popularity levels at Time 1.

Physical appearance. Consistent with our hypotheses, romantic partners' physical appearance (i.e., body appeal or physical attractiveness) did not exert a significant main effect on adolescents' peer-rated appearance, nor were there interactive effects between romantic partners' and adolescents' physical appearance in predicting Time 2 ratings of physical appearance (see Table 3).

Depressive symptoms. Romantic partners' peer-rated sadness at Time 1 was not directly associated with adolescents' Time 2 peer-rated sadness. As hypothesized, however, the interaction between adolescents' and partners' Time 1 sadness was a significant predictor (see Table 4). Results from post hoc probing of the moderating effects of adolescents' romantic partners are illustrated in Figure 4, and the results are illustrative of partner-moderated effects for other "problem-oriented" outcomes reported subsequently. Figure 4 plots the association between adolescents' Time 1 and Time 2 sadness when romantic partners' Time 1 sadness was high versus low. Adolescents' Time 1 sadness predicted their Time 2 sadness when partners' initial sadness was high ($b = 1.06$, $p < .001$) but not low ($b = -.10$, *ns*). These findings suggest that adolescents who dated partners who were initially low on sadness changed more than those who dated partners who were initially high on sadness.

Results from post hoc probing of the moderating effects of adolescents' initial sadness are illustrated in Figure 5, and the findings are illustrative of adolescent-moderated effects for subsequently reported "problem-oriented" outcomes. Here, the association between partners' sadness at Time 1 and adolescents' sadness at Time 2 is plotted when adolescents' initial sadness was high versus low. Romantic partners' Time 1 sadness predicted adolescents' Time 2 sadness, but only for adolescents who were initially high on peer-rated sadness, $b = .84$, $p < .01$ for high sad and $b = -.11$, *ns* for low sad. Among adolescents who were initially high on sadness, those who dated a partner low on sadness were less sad at Time 2 than those who dated a partner high on sadness.

Considered together, the two sets of post hoc analyses facilitate a better understanding of how the combination of adolescents' and partners' initial sadness predict change in adolescents' sadness over time. Figure 6 illustrates these patterns of change for low- and high-functioning adolescents when

Table 4
Summary of Hierarchical Regression Analyses Examining Main and Moderating Effects of Romantic Partner Characteristics on Adolescents' Depression, Aggression, and Victimization at Time 2

	PR sadness		SR depression		Relational aggression		Physical aggression		Relational victimization		Physical victimization	
	β	ΔR^2	β	ΔR^2	β	ΔR^2	β	ΔR^2	β	ΔR^2	β	ΔR^2
Step 1												
T1 adolescent (A) functioning	.40***	.05*	.49***	.54***	.12	.13***	.88***	.83***	.53***	.15***	.32**	.09**
Step 2												
T1 best friend functioning	.08	.02	.17	.04*	.32***	.09***	.08	.00	.18*	.03*	.04	.00
Step 3												
T1 romantic partner (RP) functioning	.52**	.01	.12	.01	.01	.01	−.06	.00	.11	.00	.12	.01
Step 4												
T1 A x RP functioning	.52**	.06**	.23*	.03*	.29**	.06*	−.03	.00	.25*	.03*	.07	.00
Total R^2		.14**		.62***		.29***		.83***		.21***		.10*

Note. At each step, ΔR^2 is for the step and β is for the final model. SR = self-rated; PR = peer rated.
*$p < .05$. **$p < .01$. ***$p < .001$.

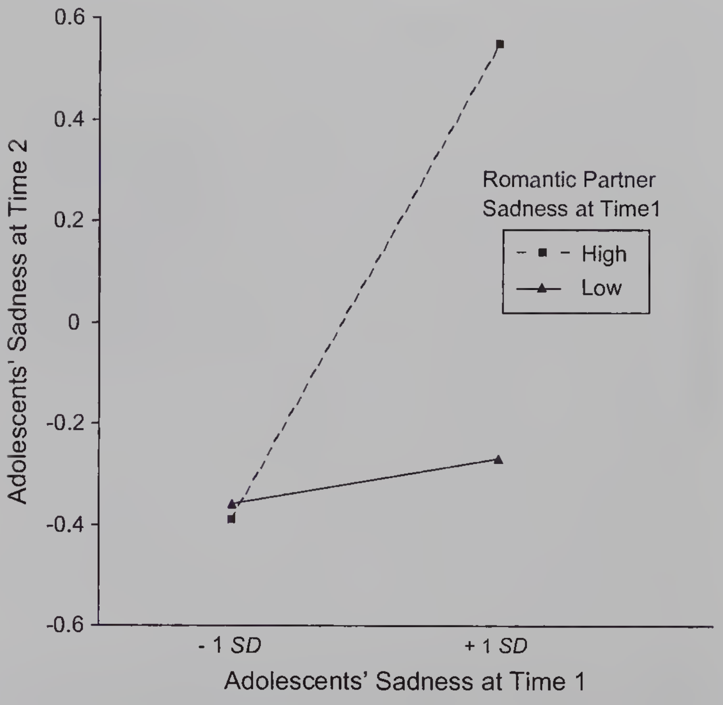


Figure 4. Graph of simple slopes for interactions between adolescent and romantic partner Time 1 peer-rated sadness in predicting adolescents' Time 2 peer-rated sadness with romantic partners' sadness as the moderator.

they dated low- versus high-functioning partners. It is also illustrative of the results for self-reported depressive symptoms, relational aggression, and relational victimization. High and low functioning are defined in relation to the sample mean ($M \pm 1$ *SD*). High-functioning adolescents (i.e., low on sadness, depressive symptoms, aggression, victimization)

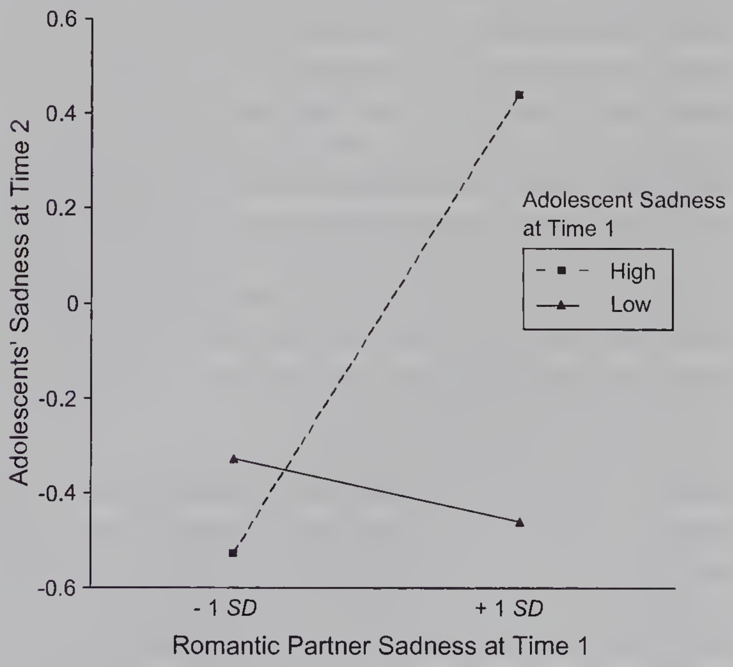


Figure 5. Graph of simple slopes for interactions between adolescent and romantic partner Time 1 peer-rated sadness in predicting adolescents' Time 2 peer-rated sadness with adolescents' sadness as the moderator.

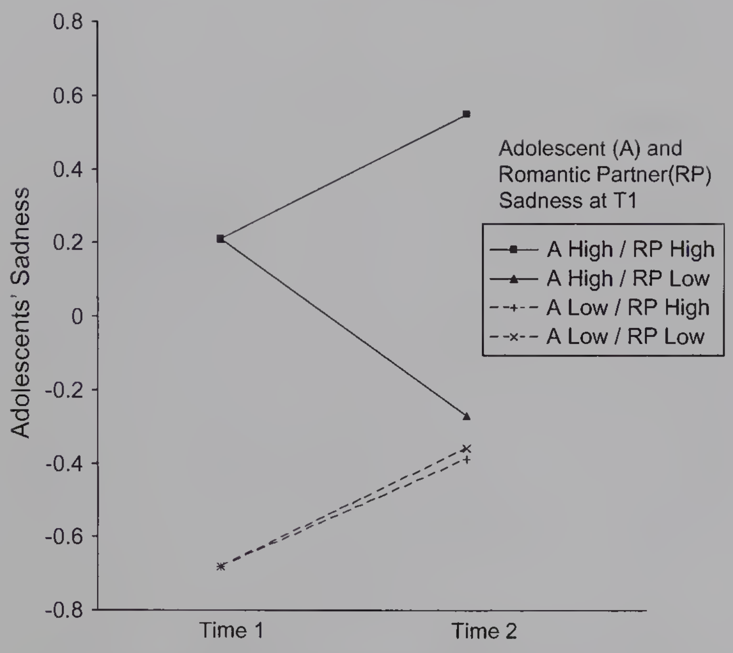


Figure 6. Graph of change patterns in adolescents' peer-rated sadness from Time 1 to Time 2 by combination of adolescent and partner sadness levels at Time 1.

remained stable over time, regardless of their partners' level of functioning. In contrast, low-functioning adolescents changed over time, but only when they dated a high-functioning partner. Low-functioning adolescents (i.e., high on sadness, depressive symptoms, aggression, victimization) who dated a high-functioning partner were indistinguishable at Time 2 from high-functioning adolescents.

For the regression predicting adolescents' self-reported depressive symptoms at Time 2, the interaction between adolescents' and partners' Time 1 depressive symptoms was again a significant predictor (see Table 4). The main effect for partners' Time 1 depressive symptoms was not significant. Results from post hoc probing of the moderating effects of romantic partners are similar to those for peer-rated sadness. Adolescents' Time 1 symptoms predicted their Time 2 symptoms when partners had high ($b = .57, p < .001$) but not low ($b = .28$, *ns*) levels of depressive symptoms at Time 1. These findings suggest that adolescents who dated partners who were low on depressive symptoms changed more than those who dated partners who were high on symptoms.

Results from post hoc probing of the moderating effects of adolescents' initial symptom levels were similar to those for peer-rated sadness. Romantic partners' Time 1 depressive symptoms predicted adolescents' Time 2 depressive symptoms, but only for adolescents who were initially high on symptoms, $b = .75, p < .05$ for high symptoms and $b = -.52$, *ns* for low symptoms. Among adolescents who were initially high on depressive symptoms, those who dated a low-symptom partner reported fewer depressive symptoms at Time 2 than those who dated a high-symptom partner.

Aggression. Romanic partners' relational aggression at Time 1 was not directly associated with adolescents' Time 2 relational aggression. However, the interaction between adolescents' and partners' Time 1 relational aggression was a significant predictor (see Table 4). Results from the post hoc probing of the moderating effects of romantic partners' Time 1 relational aggression were similar to those found for self- and peer-rated depressive symptoms. Adolescents' relational aggression at Time 1 predicted their aggression at Time 2 when partners were high but not low on relational aggression, $b = .20, p < .05$, and $b = -.15$, *ns*, respectively. These findings suggest that adolescents who dated partners who were initially low on relational aggression changed more than those who dated partners who were initially high on relational aggression.

Results from post hoc probing of the moderating effects of adolescents' Time 1 relational aggression were similar to those reported for self- and peer-rated depressive symptoms. Romantic partners' Time 1 relational aggression predicted adolescents' Time 2 relational aggression, but only for adolescents who were initially high on aggression, $b = .21, p < .05$ for high aggression and $b = -.13$, *ns* for low aggression. Among adolescents who were initially high on relational aggression, those who dated a low-aggressive partner were less aggressive at Time 2 than those who dated a high-aggressive partner.

Neither romantic partners' physical aggression at Time 1 nor the interaction between romantic partner and adolescents' physical aggression was associated with participants' Time 2 physical aggression (see Table 4).

Victimization. Although romantic partners' relational victimization at Time 1 was not directly associated with adolescents' Time 2 relational victimization, the interaction between adolescents' and partners' Time 1 relational victimization was a significant predictor (see Table 4). Adolescents' relational victimization at Time 1 predicted their victimization at Time 2 when partners were high but not low on relational victimization, $b = .36, p < .05$ and $b = -.24$, *ns*, respectively. These findings suggest that adolescents who dated partners who were initially low on relational victimization changed more than those who dated partners who were initially high on relational victimization.

Results from post hoc probing of the moderating effects of *adolescents'* Time 1 relational victimization indicated that romantic partners' Time 1 relational victimization predicted adolescents' Time 2 relational victimization, but only for adolescents who were initially high on victimization, $b = .36$, $p < .05$ for high and $b = -.22$, *ns* for low victimization. Among adolescents who were initially high on relational victimization, those who dated a partner low on victimization were less victimized at Time 2 than those who dated a partner high on victimization.

As with physical aggression, neither romantic partners' physical victimization at Time 1 nor the interaction between romantic partner and adolescents' physical victimization was associated with participants' Time 2 physical victimization after controlling for target participants' and best friends' Time 1 levels of victimization (see Table 4).

Discussion

This study is among the first to demonstrate the significance of partner selection and socialization processes in adolescents' romantic relationships. Using a longitudinal design, prerelationship similarities between young adolescent dating partners were isolated to examine patterns of partner selection. Next, we assessed romantic partner socialization by predicting changes in adolescents' psychosocial functioning over time as a function of partners' prerelationship characteristics. The findings suggest that romantic selection and socialization processes are operating even as youth are just beginning to participate in romantic activities, develop cross-sex interaction skills, and construct their identities as romantic partners (Brown, 1999; Connolly et al., 2004).

Partner Selection

Prerelationship similarities detected across both peer and self-report provide compelling evidence of young adolescents' attraction to partners who share comparable levels of social standing, appearance, and depressive symptoms. These results extend prior experimental findings that similarity is predictive of interpersonal attraction (Byrne, 1971). Similarities in peer-rated attractiveness and body appeal suggest that prior reports of similarity between partners' ratings of their own attractiveness reflect more than correspondence in self-perceptions (e.g., Feingold, 1988). Likewise, similarities in self-rated depressive features suggest that young romantic partners view themselves as comparably depressed.

The pattern of selection findings was not entirely as predicted and could underscore a distinction between two types of attraction involved in relationship formation: attraction based on similarity (Byrne, 1997; Feingold, 1988) and attraction based on broad appeal (Newcomb, Bukowski, & Pattee, 1993). Although high levels of physical attractiveness and social status have broad-based appeal, the current findings suggest that adolescents, like adults, typically pair with romantic partners who share similar levels on these dimensions (Feingold, 1988; Folkes, 1982; Hendrick & Hendrick, 1992). Adolescents also paired with partners who shared similar levels of depressive features. Whereas status and attractiveness have broad-based appeal, depressive features may only be attractive to those sharing similar characteristics (Rosenblatt & Greenberg, 1988; Wenzlaff & Prohaska, 1989). Swann and colleagues have speculated that pairing with similarly depressed partners serves to verify individuals' self-image (Giesler & Swann, 1999; Swann, 2004). The aggregation of similarly depressed youth has been noted in peer groups and close friendships during early to middle adolescence (Hogue & Steinberg, 1995; Mariano & Harton, 2005; Stevens & Prinstein, 2005) and roommate and dating relationships during late adolescence (Joiner & Katz, 1999; Katz, Beach, & Joiner, 1999). The current study extends these findings to early romantic development. Additional work is needed to pinpoint the basis of similarity. Given the parallels between young adolescents' friend and romantic relationships and the nesting of both relationships within peer groups, prerelationship similarities in depressive features could reflect peer group as well as dyadic influences.

We had expected that depressive and aggressive features would each be differentially attractive to youth and result in prerelationship similarities on both characteristics. However, we found no evidence that either physical or relational aggression led to assortative pairing. One explanation for these results is that aggression may have broad-based rather than differential appeal. Some studies suggest that aggression becomes increasingly attractive to young adolescents regardless of their own levels of aggression (Bukowski, Sippola, & Newcomb, 2000; Pelligrini & Long, 2003). Perhaps this also includes romantic attraction, in which case the general attractiveness of aggression would make selective pairing less likely to occur. The apparent absence of selective pairing on physical aggression could also reflect general developmental declines in the prominence of physical aggression in adolescent peer relationships (Nagin & Tremblay, 2005; Tremblay, Hartup, & Archer, 2005). That friend dyads in this study were similar on relational but not physical aggression supports this idea. In addition, physical aggression is more common among males and could be more appealing to females seeking male partners than vice versa (Pelligrini & Long, 2003;

Tremblay et al., 2005). Sample limitations precluded testing for gender-specific selection criteria, but their presence could have attenuated any selection effects for aggression.

Overall, the pattern of findings on partner selection suggests that similarity is important to young adolescents' romantic pairing, but only for certain characteristics, and that the pattern of attraction is similar to but not identical to that among friends. The developmental course of these selection criteria within and across relationships warrants additional attention. Patterns of romantic selection could vary according to developmental shifts in social and personal needs (Bukowski et al., 2000). Some features, such as physical attractiveness, could remain important, whereas others, such as intimacy and conflict skills, could gain significance. As the nature of adolescents' relationships change, the allure of some partner characteristics could wane in favor of others that better serve their needs.

Partner Socialization

Romantic partners' prerelationship characteristics predicted change in adolescents' functioning in various domains, including peer popularity, depression, relational aggression, and relational victimization. Of these, only popularity and depressive symptoms were important to partner selection. This pattern of findings suggests that partners need not be similar to adolescents in order to be influential.

Overall, the results from analyses examining romantic partner socialization add to the literature on peer influence in two important ways. First, they provide evidence for the significance of both friends and romantic partners to adolescent adjustment. Although best friendships were already formed at the time of our initial assessment, friend characteristics remained important predictors of adolescents' physical attractiveness, depressive symptoms, relational aggression, and relational victimization. Yet even with a conservative data analytic strategy that controlled for co-occurring best friend socialization, romantic partner characteristics emerged as significant predictors of changes in adolescents' psychosocial functioning. These findings lend credible support to theoretical assertions that romantic partners are unique and significant socializing agents (Collins, 2003).

Second, our results suggest that whether and how romantic partners affect adolescent functioning depends on characteristics of both adolescents and their partners. No partner characteristics were directly predictive of adolescents' functioning over time. This could partly reflect the stability of the characteristics assessed in this study. The small amounts of observed change in adolescents' functioning could have made it difficulty to detect any direct effects of romantic partner characteristics. Nonetheless, estimates of additive peer socialization typically range from nonsignificant to modest. Recent research on deviant peer influence suggests that these main effect models are overly simplistic, as youth vary in both their power to influence and their susceptibility to be influenced (e.g., Dishion & Dodge, 2005; Hartup, 2005). The current findings are consistent with this tenet.

Across multiple outcomes, adolescents who dated high-functioning romantic partners tended to change more than those who dated low-functioning partners, and partner characteristics were more predictive of change for low- versus high-functioning adolescents. Detecting these patterns required separate analyses of the moderating effects of adolescents' and partners' prerelationships functioning. Putting the two sets of analyses together illustrated a consistent pattern in which low-functioning adolescents appeared to benefit from dating high-functioning partners. Low-popular adolescents who dated a high-popular partner gained more popularity over time than the other three groups of adolescents. This particular finding is consistent with prior assertions that early romantic relationships serve to establish adolescents' status and popularity in the peer group (Brown, 1999). Interestingly, already popular adolescents did not seem to suffer much for dating lower popular partners, suggesting that high-popular partners could be particularly important to status grading, at least among young adolescents.

A similar pattern was found for the more problem-oriented outcomes. Only adolescents who dated high-functioning partners showed significant change over time, and romantic partner characteristics only predicted change for low-functioning adolescents. When predicting depression, sadness, relational aggression, or relational victimization, adolescents who coupled with partners who had fewer problems showed more positive changes in these areas than those who coupled with partners with more problems. Among adolescents who dated high-functioning partners, those who initially had more problems were indistinguishable at Time 2 from those who initially had few problems. These findings raise the interesting possibility that high-functioning partners could help mitigate the symptoms of more poorly adjusted youth. Similar findings have been noted among adult couples where patterns of childhood conduct disorder appear to be disrupted by the presence of supportive, nondeviant partners (Laub, Nagin, & Sampson, 1998; Quinton, Pickles, Maughan, & Rutter, 1993; Werner & Smith, 2001). In general, however, findings indicative of buffering effects are rare in the peer literature and highlight the potential for healthy peer influence at an important juncture in psychosocial development.

Additional research is needed to explain how certain adolescents with psychosocial problems pair with and benefit from high-functioning partners. It seems logical that "mismatches" would be more likely to occur on characteristics for which assortative pairing is less likely. Even within the group of characteristics for which assortative pairing is common, there could be individual differences in their relative importance, leading to less similarity on certain characteristics than others. Regardless of how mismatches occur, partner socialization could be stronger in areas that are salient to interpersonal functioning at a given age. Specifying the underlying pathways and mechanisms of these processes over the course of romantic development is an important task for future studies.

The significant stability found for adolescents who dated low-functioning partners also merits further inquiry. Within this group, the reasons for stability could vary between those who were initially high and low functioning. High-functioning adolescents could be relatively unaffected by the problems of their low-functioning partners (Adams et al., 2005). On the other hand, the stability of low-functioning adolescents with low-functioning partners echoes findings in the deviancy training literature, which indicate that psychosocial problems thrive in dyads where both youth are experiencing similar problems (Dishion, McCord, & Poulin, 1999). Our findings are consistent with these interpretations but did not directly test them. Hence, additional research is needed to examine differential reasons for stability among youth who date low-functioning partners.

This study is among the first to offer evidence that romantic partner characteristics affect the psychosocial functioning of young adolescents. The potential for partner socialization across diverse domains of functioning at this age is striking given that early romantic relationships tend to be relatively short lived and low in intimacy. Longer term studies with larger samples are needed to replicate the current findings, assess their duration, and examine potential carry over to subsequent romantic relationships. This information is critical to understanding trajectories of romantic development and their interface with psychosocial functioning. Assessing indices of positive adaptation (e.g., self-esteem, academic achievement, prosocial activity) will also be useful for understanding how romantic partners could promote development and well-being. Addressing these questions will require more detailed knowledge about the partner characteristics that are salient at different phases of romantic development. As noted earlier, there are likely to be age-related shifts in the salience of selection criteria. Similarly, the relative influence of various partner characteristics is likely to change with romantic development. For example, as romantic relationships become more intimate, partners' ways of seeking and providing support or of managing disagreements could become more potent socializing characteristics. Changes in which characteristics predict changes in functioning could also reflect shifts in adolescents' social activities, including age-related trends to engage in substance use or other health-risk behavior.

Finally, it is important to note that there are likely multiple ways in which romantic partner socialization could occur. The current study focused on estimating socialization effects within single domains, such as predicting changes in adolescents' depressive symptoms from partners' initial levels of depressive symptoms. Other, more complex processes could also be operating. For example, a given partner characteristic could influence a different adolescent characteristic. Similar findings have been noted in the adult literature, where partner support has been shown to disrupt patterns of childhood conduct disorder (Laub et al., 1998; Quinton et al., 1993). Similarly, coupling with an aggressive partner predicts increases in one's depressive symptoms (Simonelli & Ingram, 1998), and these problems could even cooccur in ways that shape both partners' behaviors (Kim & Capaldi, 2004). Such findings suggest a need for additional research to address the socialization effects of partner characteristics across domains of functioning.

Limitations and Future Directions

Several factors should be considered when interpreting the findings from this study. First, we have discussed our longitudinal findings in ways that suggest that romantic partner characteristics influence adolescents' subsequent functioning. Although our results are consistent with peer socialization processes, the data are nonexperimental and thus inconclusive with respect to causal explanations. For example, some adolescents may already be headed toward change, and their selection of certain types of partners may be a signal rather than a cause of change.

Other limitations of the current study pertain to the nature of the sample. First, only 30% of the participants from the larger data collection met our inclusion criteria of being in a romantic relationship at Time 2. This was a select group of youth, who, on average, were more physically fit, good looking, and popular, as well as less depressed and less victimized than their nondating peers. Their prevalence, however, is similar to base rate estimates of romantic relationships in this age group (e.g., Carver & Udry, 2003; Giordano, Manning, & Longmore, 2006). Although involvement in romantic relationships is not necessarily the norm for middle school students, our findings could still be relevant to the broader population of young adolescents. After all, adolescents are romantically active even when not in relationships. They are preoccupied with romantic issues and intensely aware of romantic relationship dynamics (Connolly et al., 2004).

Our sample was also restricted to middle school students who dated other middle school students at their school. Although this inclusion criterion was necessary to identify partner characteristics, the exclusion of adolescents dating peers from another school or age group could have influenced our findings. Likewise, our sample was limited to youth with same-sex friendships and other-sex romantic partners. Little is known about friend and peer group influences on romantic relationship formation in sexual minority youth. However, it seems reasonable that differences in the availability and identification of potential romantic partners for sexual minority adolescents would have implications for partner selection and socialization processes.

Overall, sample constraints in this study are balanced against numerous methodological strengths, including the assessment of diverse aspects of psychosocial functioning prior and subsequent to romantic relationship initiation, utilization of multiple reporters, isolation of friend and partner characteristics, and assessment of multiple moderators of romantic partner socialization. These assets allowed us to identify discrete patterns of selective pairing and identify ways in which the pairing of certain adolescent and partner characteristics predicts changes in adolescent adjustment. It is hoped that these findings will stimulate additional research on romantic partner characteristics, as they seem important for understanding individual differences in links between romantic involvement and adjustment.

References

Adams, R. E., Bukowski, W. M., & Bagwell, C. (2005). Stability of aggression during adolescence as moderated by reciprocated friendship status and friend's aggression. *International Journal of Behavioral Development*, *29*, 139–145.

Adams, R. E., Laursen, B., & Wilder, D. (2001). Characteristics of closeness in adolescent romantic relationships. *Journal of Adolescence*, *24*, 353–363.

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Thousand Oaks, CA: Sage.

Baker, M., Milich, R., & Manolis, M. B. (1996). Peer interactions of dysphoric adolescents. *Journal of Abnormal Child Psychology*, *24*, 241–255.

Baron, R., & Kenny, D. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173–1182.

Berndt, T. (1979). Developmental changes in conformity to peers and parents. *Developmental Psychology*, *15*, 608–616.

Blackwell, D. L., & Lichter, D. T. (2004). Homogamy among dating, cohabitating, and married couples. *Sociological Quarterly*, *45*, 719–737.

Brown, B. B. (1999). "You're going out with who?" Peer group influences on adolescent romantic relationships. In W. Furman, B. B. Brown, & C. Feiring (Eds.), *The development of romantic relationships in adolescence* (pp. 291–329). New York: Cambridge University Press.

Bukowski, W. M., Sippola, L. K., & Newcomb, A. (2000). Variations in patterns of attraction of same- and other-sex peers during early adolescence. *Developmental Psychology*, *36*, 147–154.

Byrne, D. (1971). *The attraction paradigm*. New York: Academic Press.

Byrne, D. S. (1997). An overview (and underview) of research and theory within the attraction paradigm. *Journal of Social and Personal Relationships*, *14*, 417–431.

Carver, J. R., & Udry, J. R. (2003). National estimates of adolescent romantic relationships. In P. Florsheim (Ed.), *Adolescent romantic relations and sexual behavior: Theory, research, and practical implications* (pp. 23–56). Mahwah, NJ: Erlbaum.

Cohen, J. (1988). Set correlations and contingency tables. *Applied Psychological Measurement*, *12*, 425–434.

Collins, W. A. (2003). More than myth: The developmental significance of romantic relationships during adolescence. *Journal of Research on Adolescence*, *13*, 1–24.

Connolly, J., Craig, W., Goldberg, A., & Pepler, D. (2004). Mixed-gender groups, dating, and romantic relationships in early adolescence. *Journal of Research on Adolescence*, *14*, 185–207.

Connolly, J., & Goldberg, A. (1999). Romantic relationships in adolescence: The role of friends and peers in their emergence and development. In W. Furman, B. B. Brown, & C. Feiring (Eds.), *The development of romantic relationships in adolescence* (pp. 266–290). New York: Cambridge University Press.

Dishion, T., & Dodge, K. A. (2005). Peer contagion in interventions for children and adolescents: Moving towards an understanding of the ecology and dynamics of change. *Journal of Abnormal Child Psychology*, *33*, 395–400.

Dishion, T., McCord, J., & Poulin, F. (1999). When interventions harm: Peer groups and problem behavior. *American Psychologist*, *54*, 755–764.

Dishion, T. J., & Owen, L. D. (2002). A longitudinal analysis of friendships and substance use: Bidirectional influence from adolescence to adulthood. *Developmental Psychology* , *38*(4), 480–491.

Duncan, G., Boisjoly, J., Kremer, M., Levy, D. M., & Eccles, J. (2005). Peer effects in drug use and sex among college students. *Journal of Abnormal Child Psychology*, *33*, 375–385.

Espelage, D. L., Holt, M. K., & Henkel, R. R. (2003). Examination of peer-group contextual effects on aggression during early adolescence. *Child Development*, *74*, 205–220.

Feingold, A. (1988). Matching for attractiveness in romantic partners and same-sex friends: A meta-analysis and theoretical critique. *Psychological Bulletin*, *104*, 226–235.

Folkes, V. S. (1982). Forming relationships and the matching hypothesis. *Personality and Social Psychology Bulletin*, *8*, 631–636.

Furman, W., Brown, B. B., & Feiring, C. (1999). *The development of romantic relationships in adolescence*. New York: Cambridge University Press.

Furman, W., & Buhrmester, D. (1985). Children's perceptions of the personal relationships in their social networks. *Developmental Psychology*, *21*, 1016–1024.

Furman, W., & Shaffer, L. (2003). The role of romantic relationships in adolescent development. In P. Florsheim (Ed.), *Adolescent romantic relations and sexual behavior: Theory, research, and practical implications* (pp. 3–22). Mahwah, NJ: Erlbaum.

Furman, W., Simon, V. A., Shaffer, L., & Bouchey, H. A. (2002). Adolescents' working models and styles for relationships with parents, friends, and romantic partners. *Child Development*, *73*, 241–255.

Giesler, R. B., & Swann, W. B., Jr. (1999). Striving for confirmation: The role of self-verification in depression. In T. Joiner & J. C. Coyne (Eds.), *The interactional nature of depression: Advances in interpersonal approaches* (pp. 189–217). Washington, DC: American Psychological Association.

Giordano, P. C., Manning, W. D., & Longmore, M. A. (2006). Adolescent romantic relationships: An emerging portrait of their nature and developmental significance. In A. C. Crouter & A. Booth (Eds.), *Romance and sex in adolescence and emerging adulthood: Risks and opportunities* (pp. 127–150). Mahwah, NJ: Erlbaum.

Hansen, S. L. (1977). Dating choices of high school students. *Family Coordinator*, *26*, 133–138.

Hartup, W. (1999). Constraints on peer socialization: Let me count the ways. *Merrill-Palmer Quarterly*, *45*, 172–183.

Hartup, W. (2005). Peer interaction: What causes what? *Journal of Abnormal Child Psychology*, *33*, 387–394.

Haselager, G. J., Hartup, W., van Lieshout, C. F., & Riksen-Walraven, J. M. (1998). Similarities between friends and nonfriends in middle childhood. *Child Development*, *69*, 1198–1208.

Hendrick, S. T., & Hendrick, C. (1992). *Romantic love*. Thousand Oaks, CA: Sage.

Hogue, A., & Steinberg, L. (1995). Homophily of internalized distress in adolescent peer groups. *Developmental Psychology*, *30*, 897–906.

Holmbeck, G. N. (2002). Post-hoc probing of significant moderational and mediational effects in studies of pediatric populations. *Journal of Pediatric Psychology*, *27*, 87–96.

Jaccard, J., Blanton, H., & Dodge, T. (2005). Peer influences on risk behavior: An analysis of the effects of a close friend. *Developmental Psychology*, *41*, 135–147.

Joiner, T. E., & Katz, J. (1999). Contagion of depressive symptoms and mood: Meta-analytic review and explanations from cognitive, behavioral, and interpersonal viewpoints. *Clinical Psychology: Science and Practice*, *6*, 149–164.

Kandel, D. B. (1978). Homophily, selection, and socialization in adolescent friendships. *American Journal of Sociology*, *84*, 427–436.

Kandel, D. B., Davies, M., & Baydar, N. (1990). The creation of interpersonal contexts: Homophily in dyadic relationships in adolescence and young adulthood. In L. N. Robins & M. Rutter (Eds.), *Straight and devious pathways from childhood to adulthood* (pp. 221–241). New York: Cambridge University Press.

Katz, J. W., Beach, S. R. H., & Joiner, T. E.,Jr (1999). Contagious depression in dating couples. *Journal of Social and Clinical Psychology*, *18*, 1–13.

Kim, H. K., & Capaldi, D. M. (2004). The association of antisocial behavior and depressive symptoms between partners and risk for aggression in romantic relationships. *Journal of Family Psychology*, *18*, 82–96.

Kovacs, M. (1992). *Children's Depression Inventory manual*. North Tonawanda, NY: MultiHealth Systems.

Krosnick, J. A., & Judd, C. M. (1982). Transitions in social influence at adolescence: Who induces cigarette smoking? *Developmental Psychology*, *18*, 359–368.

Laub, J. H., Nagin, D. S., & Sampson, R. J. (1998). Trajectories of change in criminal offending: Good marriages and the desistence process. *American Sociological Review*, *63*, 225–238.

Mariano, K. A., & Harton, H. (2005). Similarities in aggression, inattention/hyperactivity, depression, and anxiety in middle childhood friendships. *Journal of Social and Clinical Psychology*, *24*, 471–496.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, *27*, 415–444.

Nagin, D. S., & Tremblay, R. E. (2005). What has been learned from group-based trajectory modeling? Examples from physical aggression and other problem behaviors. *Annals of the American Academy of Political and Social Science*, *602*, 82–117.

Newcomb, A. F., Bukowski, W. M., & Pattee, L. (1993). Children's peer relations: A meta-analytic review of popular, rejected, neglected, controversial, and average sociometric status. *Psychological Bulletin*, *113*, 99–128.

Pelligrini, A., & Long, J. D. (2003). A sexual selection theory longitudinal analysis of sexual segregation and integration in early adolescence. *Journal of Experimental Child Psychology, 85*, 257–278.

Poulin, F., Dishion, T. J., & Burraston, B. (2001). 3-year iatrogenic effects associated with aggregating high-risk adolescents in cognitive-behavioral preventive interventions. *Applied Developmental Science, 5*, 214–224.

Prinstein, M. J. (2007). Moderators of peer contagion: A longitudinal examination of depression socialization between adolescents and their best friends. *Journal of Clinical Child and Adolescent Psychology, 36*, 159–170.

Quinton, D., Pickles, A., Maughan, B., & Rutter, M. (1993). Partners, peers, and pathways: Assortative pairing and continuities in conduct disorder. *Development and Psychopathology, 5*, 763–783.

Regan, P. C., & Joshi, A. (2003). Ideal partner preferences among adolescents. *Social Behavior and Personality, 31*, 13–20.

Roscoe, B., Diana, M. S., & Brooks, R. H. (1987). Early, middle, and late adolescents' views on dating and factors influencing partner selection. *Adolescence, 22*, 59–68.

Rose, A. J., Swenson, L. P., & Waller, E. M. (2004). Overt and relational aggression and perceived popularity: Developmental differences in concurrent and prospective relations. *Developmental Psychology, 40*, 378–387.

Rosenblatt, A., & Greenberg, J. (1988). Depression and interpersonal attraction: The role of perceived similarity. *Journal of Personality and Social Psychology, 55*, 112–119.

Simonelli, C. J., & Ingram, K. M. (1998). Psychological distress among men experiencing physical and emotional abuse in heterosexual dating relationships. *Journal of Interpersonal Violence, 13*, 667–681.

Skipper, J. K., & Nass, G. (1966). Dating behavior: A framework for analysis and an illustration. *Journal of Marriage and the Family, 28*, 412–420.

Steinberg, L., & Silverberg, S. B. (1986). The vicissitudes of autonomy in early adolescence. *Child Development, 57*, 841–851.

Stevens, E., & Prinstein, M. J. (2005). Peer contagion of depressogenic attributional styles among adolescents: A longitudinal study. *Journal of Abnormal Child Psychology, 33*, 25–37.

Swann, W. B. (2004). The trouble with change: Self-verification and allegiance to the self. In R. M. Kowalski & M. R. Leary (Eds.), *The interface of social and clinical psychology: Key readings* (pp. 349–356). New York: Psychology Press.

Tremblay, R. E., Hartup, W. W. & Archer, J. (Eds.). (2005). *Developmental origins of aggression*. New York: Guilford.

Urberg, K. A., Degirmencioglu, S. M., & Tolson, J. M. (1998). Adolescent friendship selection and termination: The role of similarity. *Journal of Social and Personal Relationships, 15*, 703–710.

Wenzlaff, R. M., & Prohaska, M. L. (1989). When misery prefers company: Depression, attributions, and responses to others' moods. *Journal of Experimental Social Psychology, 25*, 220–233.

Werner, E. E., & Smith, R. S. (2001). *Journeys from childhood to midlife: Risk, resilience, and recovery.* Ithaca, NY: Cornell University Press.

Child Development, November/December 2008, Volume 79, Number 6, Pages 1693 – 1706

Adrenocortical Underpinnings of Children's Psychological Reactivity to Interparental Conflict

Patrick T. Davies and Melissa L. Sturge-Apple
University of Rochester

Dante Cicchetti
University of Minnesota

E. Mark Cummings
University of Notre Dame

This study examined interrelationships among children's cortisol reactivity and their psychological reactivity to interparental conflict in a sample of 208 first graders (mean age = 6.6 years). Assessments of children's psychological reactivity to conflict distinguished among their distress, hostile, and involvement responses across multiple methods (i.e., observation, questionnaire) and informants (i.e., observer, parent). Relative to other forms of conflict reactivity, children's distress responses to interparental conflict were consistent, unique predictors of their elevated cortisol reactivity to interparental conflict even after inclusion of demographic factors as moderators and covariates. Moderator analyses further revealed that associations between distress and elevated cortisol levels in response to interparental conflict were particularly pronounced when children exhibited high levels of involvement in conflicts.

Understanding how children respond to interparental conflict is a pressing priority for public health (National Institutes of Health, 2003. Interparental conflict increases children's vulnerability to a wide range of behavioral, emotional, and social problems. In attempting to identify the processes associated with the risk experienced by children from high-conflict homes, conceptual models have underscored the significance of children's responses to conflict across multiple levels of analysis (Davies & Cummings, 2006). For example, emotional security theory (EST) regards children's elevated distress responses to conflict as developing from repeated exposure to destructive interparental conflict and, in turn, increasing the likelihood of psychological problems (Davies & Cummings, 1994). At another level of analysis, research on children's physiological reactivity to conflict has elucidated pathways between interparental conflict history and child adjustment problems (e.g., see review by Katz, 2001).

The approach commonly shared by the psychological and psychophysiological research traditions involves analyzing children's responses to conflict at one level of analysis (e.g., psychological) in isolation from children's response patterns at other levels of analysis (e.g., physiological). For example, empirical tests of psychological theories of interparental conflict have primarily focused on the analysis of behavioral and social-cognitive reactions to interparental conflict to the relative exclusion of physiological assessments (Davies & Cummings, 1994; Grych & Fincham, 1990). Likewise, studies conducted on children's physiological functioning in the context of adult or parental conflict are commonly rooted in a different body of literature addressing children's physiological reactivity to broad array of stressful events (Katz, 2001). Therefore, little is known about the interplay among children's psychological and physiological reactivity to interparental conflict.

In reflecting the repeated calls to integrate children's physiological and psychological responses to family stress (Cummings, Davies, & Campbell, 2000; Fox, Hane, & Perez-Edgar, 2006; Katz, 2001), Katz concluded: "It is only through integrating our understanding of both biological and behavioral processes that we can get a complete picture of the effects of marital conflict on the whole child" (p. 207). Therefore, although dissociations between physiological and

---

This research was supported by the National Institute of Mental Health (Project R01 MH57318) awarded to P.T.D. and E.M.C. and the Spunk Fund, awarded to D.C. The authors are grateful to the children, parents, teachers, and school administrators who participated in this project. Their gratitude is also expressed to project staff, including: Courtney Forbes, Courtney Henry, Marcie Goeke-Morey, Amy Keller, Michelle Sutton, Alice Schermerhorn, and the graduate and undergraduate students at the University of Rochester and University of Notre Dame.

Correspondence concerning this article should be addressed to Patrick T. Davies, Department of Clinical and Social Sciences in Psychology, University of Rochester, Rochester, NY 14627. Electronic mail may be sent to patrick.davies@rochester.edu.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7906-0011

psychological indices of functioning are common in the literature, fully understanding children's responses to conflict at multiple levels of analysis requires further examination of how indices of physiological reactivity to conflict correspond concurrently with comparable assessments of psychological responses to conflict (Fox et al., 2006).

To address the paucity of research on the relationship between children's physiological and psychological reactivity to interparental conflict, we examine associations between children's behavioral reactivity and their limbic-hypothalamic-pituitary-adrenocortical (LHPA) system reactivity to interparental conflict. Through its role as a second wave of autonomic responding to stress (Cahill & McGaugh, 1998), the LHPA axis is regarded as a useful means of advancing an understanding of children's reactivity to interparental conflict (Saltzman, Holden, & Holahan, 2005). In response to stressful events, components of the limbic system (e.g., amygdala, hippocampus) involved in processing aversive stimuli modulate the release of corticotropin-releasing factor (CRF) by the hypothalamus. CRF, in turn, activates the adrenal gland to secrete cortisol by stimulating the pituitary gland to produce and release adrenocorticotropic hormone into the bloodstream. Increases in cortisol in response to stress mobilize energy (e.g., glucose, oxygen), increase cardiovascular activity, and modulate the processing, learning, and memory consolidation of emotionally significant events (Cahill & McGaugh, 1998; Gunnar & Vazquez, 2006). In accord with the role that the LHPA axis plays in allocating resources to process and cope with stressful events, cortisol reactivity may be a critical component of children's responses to the stress of witnessing interparental conflict. Thus, the purpose of this study was to examine the concomitant correspondence between children's cortisol reactivity to interparental conflict and three of the most common forms of children's responses to conflict: distress, hostility, and involvement (Cummings & Davies, 1994).

EST provides a foundation for expecting associations between children's behavioral and physiological responses to interparental conflict. According to EST, preserving a sense of security is an important goal for children in contexts of interparental conflict (Cummings & Davies, 1996). When children witness interparental conflict, elevations in their distress characterized by anxiety, vigilance, and emotional upset are theorized to be one of the most reliable barometers of children's insecurity in the interparental relationship. Drawing on the role of the LHPA axis in marshalling resources to cope with threatening events, EST proposes that distress indicative of children's concerns about security may be linked with changes in the ability of the LHPA axis to mount physiological responses to interparental conflict. These individual differences in the LHPA axis are hypothesized to insidiously undermine children's physical and mental health over time (Davies, Winter, & Cicchetti, 2006).

Because EST offers little guidance with regard to the nature of the proposed link between behavioral and cortisol reactivity, theories on the physiological underpinnings of individual psychological functioning may sharpen predictions on the interplay between children's psychological and cortisol reactivity to interparental conflict. First, consistent with the emphasis on distress in EST, some physiological conceptualizations postulate that elevated cortisol reactivity to stressful events is most likely to occur when children are predisposed toward exhibiting submissive behavioral reactions characterized by high levels of emotional distress, anxiety, and inhibition (e.g., Fox et al., 2006; Kagan, Reznick, & Snidman, 1998; Spangler & Grossmann, 1993). Thus, in the context of understanding children's responses to interparental conflict, this "distress specificity" model would predict that children's distress should evidence unique and specific associations with cortisol reactivity relative to other forms of behavioral responding (i.e., involvement, hostility). Second, the negative affect hypothesis postulates that greater cortisol reactivity underlies more general forms of negative emotional responses to stressful events, including irritability and hostility (Spangler & Grossmann, 1993). Therefore, in contrast to the distress specificity model, the prediction of the negative affect hypothesis is that both distress and hostility would evidence stronger relations to elevated cortisol reactivity, especially in relation to involvement. Third, within a general arousal model, any type of behavioral activation or engagement with the stressor is expected to be associated with elevated cortisol reactivity (see Gunnar & Vazquez, 2006). Thus, as forms of behavioral activation, another hypothesis is that distress, hostility, and involvement may each be linked with greater cortisol reactivity.

Some support is evident for the predictions of the distress specificity model. Children with elevated internalizing symptoms have been shown to exhibit concurrent elevations in cortisol reactivity to mild laboratory stressors (e.g., Ashman, Dawson, Panagiotides, Yamada, & Wilkinson, 2002; Granger, Weisz, & Kauneckis, 1994). Tests of the coupling of behavioral and adrenocortical activity in response to the same stressor also provide support to the distress specificity hypothesis. For example, children's behavioral distress in response to social stressors (e.g., caregiver separation, transition to day care) have been contemporaneously linked with greater cortisol reactivity to social stressors (e.g., caregiver separation, transition to day care) (Dettling, Gunnar, & Donzella, 1999; Gunnar, Larson, Hertsgaard, Harris, & Brodersen, 1992). Likewise, patterns of distress reflecting insecure attachment patterns are associated with children's greater cortisol reactivity to the series of separations and reunions during the strange situation (Hertsgaard, Gunnar, Erickson, & Nachmias, 1995; Spangler & Grossmann, 1993).

At the same time, other studies do not definitively support the distress specificity hypothesis. Inclusion of indicators of anger in assessments of distress in some of the studies described above cannot disentangle the distinct associations between forms of negative emotional and adrenocortical reactivity as outlined in the distress specificity hypothesis (e.g., Dettling et al., 1999). In fact, some studies have revealed that indices of child aggression, anger, and global negative affect in child care settings were correlated with increases in cortisol activity across the day (Dettling, Parker, Lane, Sebanc, & Gunnar, 2000; Dettling et al., 1999; Gunnar & Donzella, 2000; Tout, de Haan, Campbell, & Gunnar, 1998). Therefore, in support of the negative affect model, one alternative hypothesis would be that high levels of cortisol reactivity are coupled with any form of negative reactivity to interparental conflict. Given that some of the studies on preschool adjustment to day care have also reported that impulsive and approach behaviors are associated with elevated child cortisol reactivity (e.g., Dettling et al., 1999), the general arousal hypothesis that any effortful response to interparental conflict may covary with elevated cortisol is also plausible.

In light of inconsistencies in associations between behavioral and adrenocortical reactivity to other stressful events, rigorously testing the distress specificity hypothesis hinges on examining its relative viability in relation to the negative affect and general arousal hypotheses.

Thus, the present study simultaneously examines the specificity and commonality of patterns of associations between children's distress, hostile, and involvement responses to interparental conflict and elevated cortisol reactivity to interparental conflict. Consistent with the organizational perspective on development (Cicchetti & Tucker, 1994), the associations between any form of behavioral responding and adrenocortical reactivity may depend on children's broader patterns of behavioral responses to interparental conflict. For example, the EST suggests that distress responses may signify more pronounced threats to children's security if they are coupled with high levels of involvement in interparental conflicts (Davies et al., 2006). Accordingly, we also employ moderator models to explore whether associations between distress and cortisol reactivity are moderated by involvement and hostility. Toward the goal of examining the stability of our results, a final step is to examine the role of child and demographic characteristics (i.e., gender, race, blood contamination in saliva samples, medication use, research site) that may alter associations between children's psychological and cortisol reactivity to conflict as covariates and moderators (Cohen et al., 2006; Granger et al., 2007; Hibel, Granger, Cicchetti, & Rogosch, 2007; Loney, Butler, Lima, Counts, & Eckel, 2006).

Method

Participants

The data for this study were collected from the second measurement occasion of a longitudinal project on children's coping with family stress. The original sample of 235 kindergarten children and their families in the first measurement occasion of the study were recruited through local school districts and community centers at two research sites located in moderate-sized cities in the Northeast and the Midwest. The sample for this study consisted of 208 mothers and their first-grade children. Reductions in the size of the original sample resulted from attrition of families during the 1-year span between measurement occasions ($n = 8$), incomplete data across the three cortisol assessments ($n = 2$), and exclusion of children who were identified as outliers (i.e., ± 3 *SD* from the mean) on cortisol assessments ($n = 8$). Due to the potential effects of psychotropic and glucocorticoid drugs on cortisol levels (Hibel et al., 2007; Masharani et al., 2005), 9 additional children who were taking these medications were excluded from the final sample. Thus, the sample used in this article consisted of 208 mothers and their children. Statistical comparisons of the present sample ($n = 208$) and participants in the larger sample who were excluded from this study ($n = 27$) along 16 demographic (e.g., family income, occupational status, race/ethnicity, marital status, age) variables yielded no significant differences.

Sociodemographic data reflected that the participating families were similar to the households in the counties from which our sample was drawn (i.e., St. Joseph County, IN; Monroe County, NY). Median annual family income of the families was between $40,000 and $54,999. Mothers, on average, completed some college. The majority of the sample of mothers and children were White (75%), followed by smaller proportions of Black (15%), multiracial (7%), and Hispanic (3%) family members. A large percentage of female caregivers reported being the biological parents of their children (95%), followed by relatively smaller percentages of step or adoptive parents (3%), and other types of guardianship (2%). The mean age of the children was 6.6 years (*SD* = 0.52), with 56% of the sample consisting of girls (n = 117) and 44% consisting of boys (n = 91). Mothers, fathers, and children lived in the same household for an average of 6.4 years (*SD* = 1.09).

Procedures

Following the completion of two visits to the laboratory during the first measurement occasion, mothers and children attended two more laboratory visits at one of the research sites during second measurement occasion 1 year later. Data for the current study were drawn from the second visit of the second measurement occasion (i.e., the fourth family visit to our laboratory). The laboratories at each site were designed to be comparable in size and quality and included: (a) an observation room that was designed to resemble a family room (e.g., couch, pictures, lamps, end tables) and equipped with audiovisual equipment to capture family interactions and (b) interview rooms for completing confidential survey measures. The research was approved by the Institutional Review Boards at both sites prior to conducting the study.

Simulated phone argument task (SPAT). During the visit, children and their mothers participated in the SPAT to assess child reactivity to interparental conflict. During this procedure, children witnessed a live simulated conflict and resolution between their parents over the telephone. Each exchange lasted approximately 1 min and was interspersed by a 3-min free period to assess prolonged or delayed responses of children. The conflict script revolved around a relatively trivial disagreement regarding whether the father had completed a task requested by the mother (i.e., stopped at the store or made a phone call or an appointment). The mothers were instructed to convey mild irritation, frustration, and anger toward their partner as they normally would at home. Although the simulation indicated to the child that the father was on the other end of the phone, an experimenter was actually on the phone feeding the mother the lines from the script. Videotaped records were later coded for child overt reactivity to the interparental conflict. For ethical purposes of allaying any child distress, the resolution with the father over the telephone consisted of mother communicating a moderate level of understanding, caring, and warmth in her tone of voice. Interspersed between the free period following the conflict and the resolution were two 5-min interviews with children designed to assess their reactions to the conflict and resolution, respectively.

Several procedures were instituted during a presimulation training session to maximize the validity of the SPAT. First, to maximize the accuracy of the simulations, mothers listened to a standard, audiotaped sample of the conflict and resolution and practiced the script with the experimenter until they were able to convey accurately the content and affective tone of the exchanges. Moreover, in feeding the mother the lines during the procedure, the experimenter simulated the affective tone and level for the mother to emulate. Second, to increase the ecological validity of the exchanges, mothers: (a) scheduled the visit during a time in which the father was accessible by phone, (b) chose one of the three script options that most realistically reflected the nature of interparental activities, and (c) had the opportunity to alter the scripts to reflect the content and nature of actual interparental conflicts and resolutions that occur in the home. Script changes were only implemented if they did not significantly alter the intensity, meaning, and affective tone of the disagreement or resolution. Thus, this procedure provided important controls over the nature of interparental conflict stimuli while at the same time presenting conflicts that were for all purposes live and real from the observing child's perspective. The validity of the SPAT is supported by associations between children's distress reactions to the simulated conflict and their exposure to family conflict and their concurrent and prospective psychological problems (e.g., Davies, Cummings, & Winter, 2004).

Salivary cortisol collection. Saliva samples were collected from the children at three points during the simulated conflict procedure to obtain cortisol measures. Cortisol levels tend to evidence more gradual declines during the afternoon and evening hours than during the morning hours (Knutson et al., 1997; Stansbury & Gunnar, 1994). Thus, to limit the effects of time of day on cortisol assessments, sample collection times for cortisol were, with one exception, collected in the afternoon or early evening. Average sampling time for pretask cortisol occurred at 3:50 p.m. (*SD* = 2 hr 1 min).

In the initial 20 min of the visit prior to the collection of the pretask saliva sample, the experimenters developed rapport with the families, obtained parental consent and child assent, and invited children to play with toys to get reacquainted with the laboratory. While the mother was learning the script for the SPAT, the children then followed conventional sampling procedures in preparation of saliva sampling procedures (Schwartz, Granger, Susman, Gunnar, & Laird, 1998). Children rinsed their mouths with water prior to the baseline assessment to limit the undue influence of various contaminants during the assay process. Following a 7-min period in which they were invited to play with toys, children chewed Trident original flavor sugarless gum for approximately 3 min to stimulate saliva flow immediately prior to saliva collection. The pretask sample was then collected prior to the simulated conflict by having children expurgate through a plastic straw directly into a 20-ml collection vial.

Two postconflict saliva samples were also obtained to assess trajectories of cortisol change across three assessments. No definitive guidelines are available for precisely identifying the timing of peak cortisol levels following stressors. However, a recent meta-analysis of cortisol functioning revealed that cortisol levels across 10-min periods following the stressor were highest during the 21- to 30-min and 31- to 40-min epochs than any other 10-min period (Dickerson & Kemeny, 2004). Therefore, the two postconflict saliva samples were obtained approximately 25 and 35 min after the simulated conflict to correspond with the midpoints of the two 10-min peak periods of cortisol reactivity to stressors. Consistent with the preconflict sampling procedures, children chewed Trident original flavor sugarless gum to stimulate saliva flow and then expectorated through a plastic straw directly into a 20-ml collection vial. In the interim between the SPAT and the two postconflict sampling procedures, children were invited to play with toys in the room to limit extraneous stress and challenge that may result from inclusion of tasks or measures. The three saliva samples were immediately stored at −36°C until it was shipped on dry ice to Salimetrics LC (State College, PA).

Reports of child reactivity to conflict. Following a brief break after the SPAT, mothers completed a measure of children's behavioral reactivity to interparental conflict in the home as part of a battery of questionnaires.

Measures

Observer ratings of child behavioral reactivity to conflict. Children's behavioral reactions to the simulated conflict during SPAT were coded for distress, hostility, and involvement. To assess children's behavioral reactivity to interparental discord, judges coded the presence of specific types of the three behaviors (i.e., distress, hostility, involvement) for eight 30-s intervals during the 4-min videotaped segment of the conflict and postconflict periods. Specific behaviors indicative of distress reactions included: (a) freezing—tense, motionless, or "fixed in place" for 5 or more seconds; (b) anxiety—expressions of anxiety through facial expressions and/or physical gestures; and (c) verbal concern—verbalizations of concern about the interparental conflict. Measures of hostile behavior consisted of: (a) anger—facial expressions or postures reflecting anger, (b) interpersonal aggression—verbal or physical hostility directed toward the self or others, and (c) object aggression—verbal or physical aggression directed toward objects. Finally, specific indicators of involvement were as follows: (a) comforting the parent—verbal or physical attempts to comfort or mediate the conflict, (b) inquiries about parent feelings or relationships—questions about the emotional state of the parent or quality of the interparental relationship (e.g., "Mom, are you okay?" "Is dad mad?"), and (c) triangulation—intervention strategies that reflect efforts to ally with one parent against the other parent. Although codes within each composite were designed to economically capture complementary dimensions (i.e., distinct with modest overlap) of the larger construct, codes were not mutually exclusive. Therefore, multiple behavioral codes (e.g., anger, interpersonal aggression) could be rated as present within any interval. Frequency counts of the specific indicators of each of the three respective constructs were summed together to form composites of distress, hostile, and involvement behaviors. To establish reliability, two coders independently coded 25% of the videotapes. Intraclass correlation coefficients, reflecting interrater agreement in the frequency counts of each of the nine specific behavioral codes ranged from .64 to 1.0. Average intraclass coefficients were .86 for indicators of distress, .74 for the hostility codes, and .91 for the involvement behaviors.

Maternal reports of behavioral reactivity to conflict. Mothers completed questionnaire items derived from the Security in the Marital Subsystem – Parent Report Inventory (SIMS – PR) to obtain indices of children's distress, hostile, and involvement responses to conflict in the home over the last year that were comparable to behavioral codes in the SPAT (Davies, Forman, Rasi, & Stevens, 2002). For the current report, three modified versions of the SIMS – PR subscales were utilized. The Distress (or Emotional Reactivity) scale consists of six items that reflect anxiety, distress, and submissiveness (i.e., "appears frightened," "keeps very still [almost as if he or she is frozen]," "still seems upset after we argue," "appears upset," "tries to hide feelings," "repeatedly brings up questions and concerns about arguments after its over"). The Hostility scale contains five items that characterize angry and aggressive reactions to interparental conflict in the home (i.e., "yells at family members," "says unkind things to family members," "starts hitting, kicking, slapping, or throwing things at family members," "appears angry," "causes trouble"). The Involvement scale is composed of five items that index children's behavioral intervention in interparental conflicts and their aftermath (e.g., "tries to help us solve the problem," "tries to comfort one or both of us," "watches and listens very closely," "tells us to stop arguing," "gets involved in the argument"). In further support of the factor structure of the SIMS – PR, an exploratory factor analysis of the 16 items generated a three-factor solution that supported the specification of each of the items onto three respective subscales. Internal consistencies for the Distress, Hostility, and Involvement scales, as indexed by alpha coefficients, were .78, .78, and .73, respectively. The validity of the original SIMS – PR scales is supported by its theoretically meaningful associations with children's exposure to family conflict, coping difficulties with family stress, and psychological problems (e.g., Cummings, Schermerhorn, Davies, Goeke-Morey, & Cummings, 2006; Davies, Harold, Goeke-Morey, & Cummings, 2002).

Cortisol. All samples were assayed for salivary cortisol at Salimetrics Inc. in duplicate using a highly sensitive enzyme immunoassay (Salimetrics, PA). The test used 25 μl of saliva per determination, has a lower limit of sensitivity of 0.003 μg/dl, standard curve range of 0.007 – 1.8 μg/dl, and average intra- and interassay coefficients of variation 5.1% and 8.2%, respectively. Method accuracy, determined by spike recovery, and linearity, determined by serial dilution are 103% and 96%, respectively. Values from matched serum and saliva samples show the expected strong linear relationship, $r(63) = 0.89$, $p < .0001$ (Salimetrics, 2000).

Demographic and child characteristics. Five demographic and child variables were examined as potential covariates or moderators in the model: (a) blood contamination in saliva, (b) child medication use, (c) research site, (d) child gender, and (e) child race. First, to obtain an assessment of blood contamination, saliva samples were assayed for transferrin through immunoassays conducted by Salimetrics. The test, which uses 20 μl of saliva per assay, has a minimum detection limit of 0.12 mg/dl. Mean blood contamination levels (M = 0.33 mg/dl; SD = 0.35 mg/dl; range = .05 – 3.22 mg/dl) in this study were generally comparable to, if not lower than, prior studies (Kivlighan et al., 2004). Second, consistent with prior work (e.g., Hibel, Granger, Kivlighan, Blair, & the Family Life Investigators, 2006), mothers reported on children's over-the-counter and prescription medication use. Children who were taking at least one medication (20%) were compared with children who were medication free (80%). The most frequently used medications included acetaminophen or ibuprofen (6.7%) and nonsteroidal anti-inflammatory drugs (6.7%), followed by relatively lower levels of over-the-counter cough/cold medicine (2.4%), vitamins (2.4%), antibiotics (1.4%), antacids (0.9%), and other medications (2.9%). Third, research site was coded as a dichotomous variable (i.e., Northeast site contrasted with Midwest site). Finally, in addition to examining child gender (i.e., girls contrasted with boys), child race was explored as a fifth potential variable based on prior LHPA axis differences between White and Black adults (Black children contrasted with White children; Cohen et al., 2006).

Results

Descriptive Analyses

Table 1 provides the raw means, standard deviations, and correlations among the focal variables in the primary analyses. Correlations between distress, hostility, and involvement within each method of assessment (i.e., SPAT observation, maternal report) were modest in magnitude (mean r = .20, range = −.03 to .35). As expected, associations among conceptually related constructs (e.g., involvement) derived from different methods were also modest and significant (mean r = .20, range = .15 to .24). Modest correspondence across different informant reports of child psychological distress is very common in developmental and clinical research (De Los Reyes & Kazdin, 2005). As with prior research on child psychological problems, the multiple methods were derived from informants who amassed different types of information about children from different contexts (Achenbach, 2006). Log transformations of the three cortisol measures were used to reduce skewness and normalize distributions prior to the primary analyses, yielding the following means and standard deviations: preconflict (M = −2.44, SD = 0.47), post-I conflict (M = −2.70, SD = 0.44), and post-II conflict (M = −2.75, SD = 0.43).

Table 1

Means, Standard Deviations, and Intercorrelations of the Primary Variables in the Study

Variables	*M*	*SD*	1	2	3	4	5	6	7	8
Child interparental conflict reactivity (simulated phone argument task observation)										
1. Distress	1.32	1.39	—							
2. Hostility	0.39	1.04	−.03	—						
3. Involvement	0.45	0.71	.28*	−.02	—					
Child interparental conflict reactivity in the home (maternal report)										
4. Distress	9.75	3.84	.15*	.07	.07	—				
5. Hostility	7.00	2.74	.04	.24*	−.02	.34*	—			
6. Involvement	12.40	4.45	.09	−.05	.20*	.35*	.25*	—		
Child cortisol reactivity										
7. Preconflict (μg/dl)	0.097	0.044	−.11	.05	.00	−.06	.03	.02	—	
8. Post I (μg/dl)	0.073	0.032	.01	.00	.05	.04	−.03	.03	.71*	—
9. Post II (μg/dl)	0.070	0.030	.01	.03	−.01	−.03	−.03	−.03	.60*	.86*

Note. $*p < .05$.

Preliminary Analyses

The predominant strategy of assessing cortisol reactivity to stressors through the piecewise calculation of difference scores between prestressor and each poststressor cortisol level in analysis of variance or regression models fails to: (a) disentangle measurement error from measures of reactivity and (b) effectively capture individual differences in cortisol trajectories across three or more measurement occasions (Hruschka, Kohrt, & Worthman, 2005). Because growth modeling addresses these limitations (e.g., see Powers, Pietromonaco, Gunlicks, & Sayer, 2006), we assessed the slope of change across the preconflict, post-I, and post-II cortisol assessments through the use of latent growth curve (LGC) modeling.

Given the diurnal rhythm of cortisol, our first step in LGC modeling involved specifying the effects of time of measurement, prior to the inclusion of the main predictors or any covariates or moderators. LGC specification was identical to the model depicted in Figure 1 without the inclusion of the three forms of conflict reactivity as predictors. The model fit the data well, $\chi^2(7, N = 208) = 2.84$, $p = .90$, χ^2/df ratio = 0.41, root mean square error of approximation (RMSEA) = 0.000, comparative fit index (CFI) = 1.00. As an estimate of preconflict cortisol levels, the mean level of the intercept was significantly different from 0 ($\mu_i = -1.60$, $z = 6.37$). The mean slope parameter, which characterizes average, constant change in cortisol across the three assessments, was positive but not significant ($\mu_s = 0.44$, $z = 1.67$). Notably, analyses of the variance for the intercept and slope factors yielded significant individual differences in the initial levels of cortisol ($\zeta_i = 0.18$, $z = 6.60$) and change in cortisol ($\zeta_s = 0.18$, $z = 4.23$). Therefore, the results support our aim of examining associations between individual differences in children's psychological and cortisol responses to interparental conflict.

Primary Analyses

Although associations between behavioral and adrenocortical functioning may reflect bidirectional influences between the systems, our goal was to identify the specificity and uniqueness of associations between children's patterns of behavioral reactivity and their cortisol reactivity. Accordingly, to examine the relative and distinct patterns of correspondence between behavioral and cortisol reactivity, we simultaneously examined children's distress, hostility, and involvement as predictors of cortisol activity in a series of growth models. The first model involved the specification of pathways from SPAT observational ratings of distress, hostility, and involvement to the initial levels (intercept) and change in cortisol in response to the conflict (slope). The model, which is depicted in Figure 1, provided an acceptable representation of the data, $\chi^2(19, N = 208) = 15.34$, $p = .70$, χ^2/df ratio = 0.81, RMSEA = 0.000, CFI = 1.00. Paths between children's behavioral responses and the cortisol slope are of particular relevance in light of our focus on children's cortisol reactivity to conflict. Consistent with hypotheses, children's distress was positively associated with cortisol reactivity to the simulated interparental conflict, $\beta = .23$, $p < .01$. In contrast, hostility and involvement were not associated with initial levels or change in cortisol reactivity.

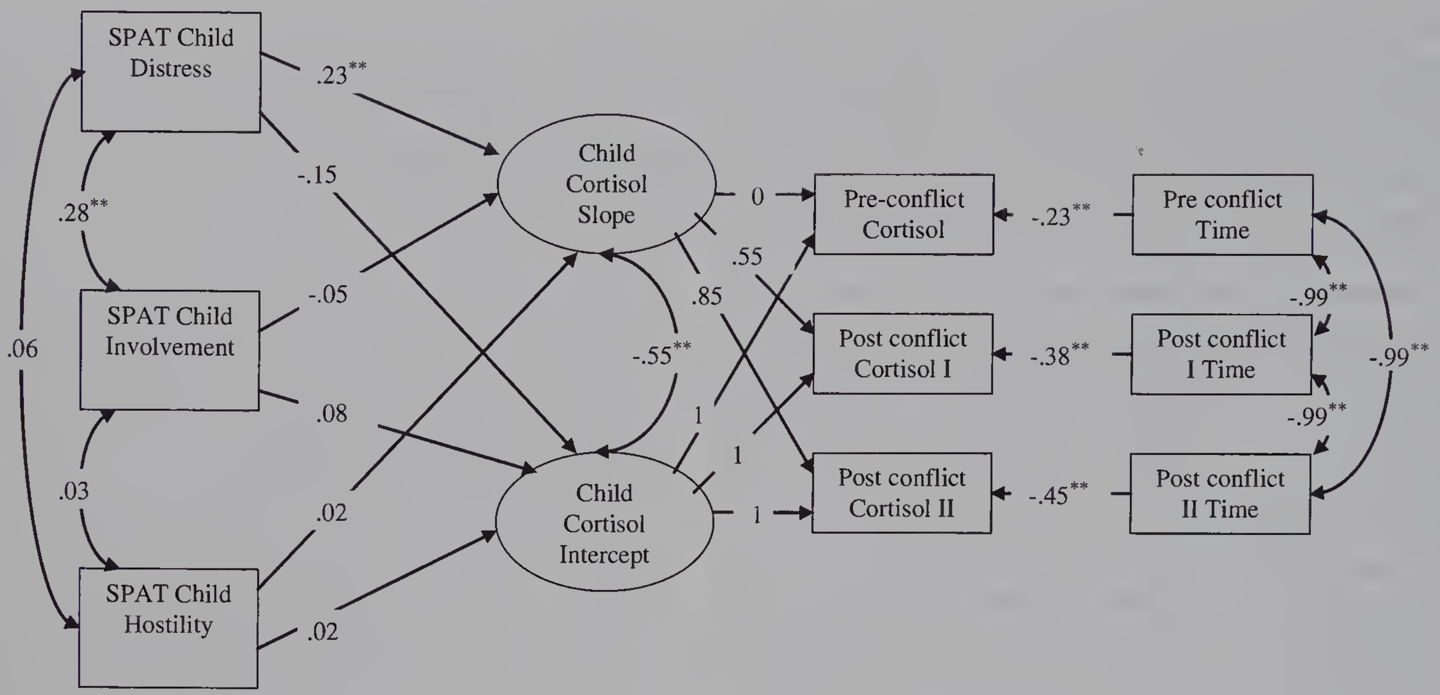


Figure 1. Associations between children's observed psychological reactivity and their cortisol reactivity to the simulated interparental conflict.
Note. SPAT = simulated phone argument task.
$^{**}p < .01$.

As a further test of the specificity of associations, we calculated pairwise parameter comparisons of the magnitude of paths running from each form of child psychological reactivity to child cortisol reactivity. The results indicated that the association between distress and cortisol reactivity was marginally stronger than comparable pathways involving hostility, $z = 1.60$, $p = .055$ (one-tailed), and involvement, $z = 1.60$, $p = .055$ (one-tailed).

To explore the viability of this model for children's psychological reactivity in the home, we also examined associations between maternal reports of children's reactivity to conflict in the home and their cortisol reactivity to the simulated interparental conflict. As shown in Figure 2, the model yielded a good representation of the data, $\chi^2(19, N = 208) = 34.16$, $p = .02$, χ^2/df ratio = 1.80, RMSEA = 0.062, CFI = 0.99. Consistent with the findings from the first model, maternal reports of children's distress responses to interparental conflict in the home were significantly associated with greater cortisol reactivity to conflict, $\beta = .27$, $p < .001$, and lower preconflict levels of cortisol, $\beta = -.18$, $p < .05$. Conversely, dampened cortisol reactivity was related to maternal reports of child hostility, $\beta = -.20$, $p = .01$. Analysis of pairwise comparisons among the pathways running from behavioral reactivity to the cortiso l slope revealed that distress was a stronger predictor of greater cortisol reactivity than either hostility, $z = 3.53$, $p < .01$, or involvement, $z = 3.08$, $p < .01$.

Tests of the Stability of the Primary Findings

As a further test of the generalizability of the results, we examined whether blood contamination in saliva, child medication use, research site, child gender, and child race may serve as both covariates and moderators that alter the nature of associations between children's distress and their cortisol reactivity to conflict. First, for both the SPAT and the maternal report models of child reactivity depicted in Figures 1 and 2, we conducted five additional covariate analyses incorporating paths between each of the five demographic variables and the intercept and slope of child cortisol. Supporting the stability of the findings, all paths between the covariates and the intercept and slope of cortisol were nonsignificant. Moreover, inclusion of the demographic variables did not alter the significant paths between children's distress and child cortisol reactivity for either the SPAT or the maternal ratings of children's responses to conflict. Second, moderator models examined whether the associations between distress and cortisol varied across the five demographic variables. We conducted separate models for each predictor and demographic

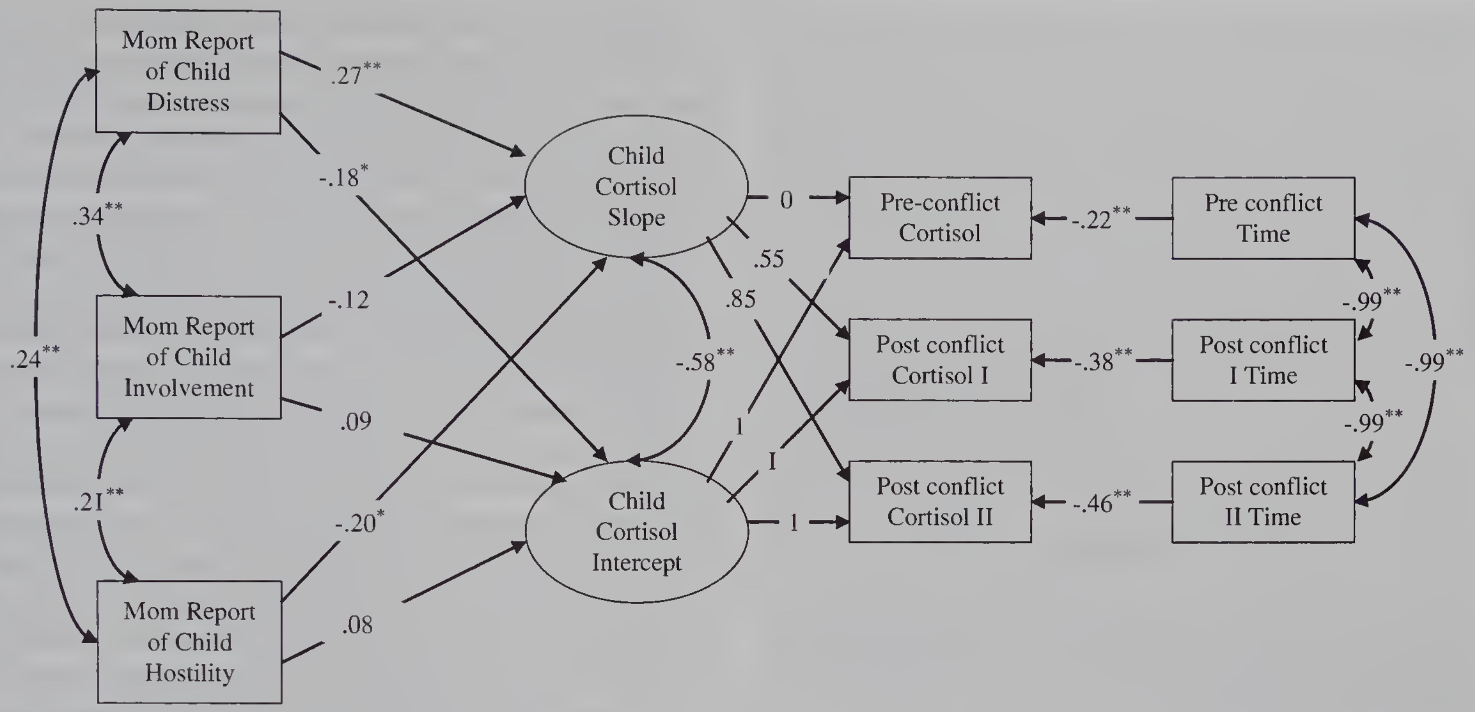


Figure 2. Associations between maternal reports of children's psychological reactivity to interparental conflict in the home and their cortisol reactivity to the simulated interparental conflict.
$^{*}p < .05$. $^{**}p < .01$.

variable pairing. In these analyses, the main effects of the distress variable and the specific demographic factor along with the two-way interaction term were all included as predictors of the cortisol intercept and slope. The results of the analyses produced no more moderating effects for the five demographic variables (1 of the 20 interactions was significant) than would be expected by chance. In conclusion, the results were consistent across a variety of contextual conditions, including research site, child gender, child race, blood contamination levels, and child medication use.

Child Involvement and Hostility as Moderators of Distress

Guided by our conceptual focus on distress, our final analytic model selectively examined whether associations between distress and child cortisol reactivity to conflict were moderated by children's involvement or hostile reactions to the conflict. Distress, proposed moderator (i.e., either hostility or involvement), and the Distress × Proposed Moderator (i.e., either hostility or involvement) interaction variables were simultaneously estimated as predictors of the cortisol intercept and slope in the LGC analyses. Four separate statistical models were specified for each of the two measures (SPAT, maternal ratings) of the two proposed moderators after centering the predictors to reduce problems with multicollinearity (Aiken & West, 1991). For the sake of brevity, we focus on the results that directly address the paths between moderator terms and cortisol reactivity. Of the four moderator analyses conducted, only the interaction term involving SPAT observations of distress and involvement was a significant predictor of child cortisol reactivity ($B = .07$, $t = 2.38$, $p < .05$). To clarify the interaction in LGC analyses of the SPAT, we followed the procedures outlined by Preacher, Curran, and Bauer (2006) for: (a) plotting regression slopes of child distress at high (1 *SD* above the mean) and low (1 *SD* below the mean) levels of involvement and (b) conducting simple slope analyses to examine whether associations between distress and cortisol reactivity were significantly different from 0 at high (+1 *SD*) and low (−1 *SD*) levels of involvement. As the graphical plot in Figure 3 shows, the association between distress and the cortisol slope is particularly pronounced at high rather than low levels of involvement. In support of this interpretation, simple slope analyses indicated that observations of child distress predicted greater cortisol reactivity when levels of child involvement in the conflict were high ($B = .15$, $t = 4.24$, $p < .001$) but not low ($B = -.01$, $t = 0.25$, $p = ns$).

Discussion

Various conceptualizations of children's coping with interparental conflict and children's psychobiological reactivity to social stressors converge to highlight the

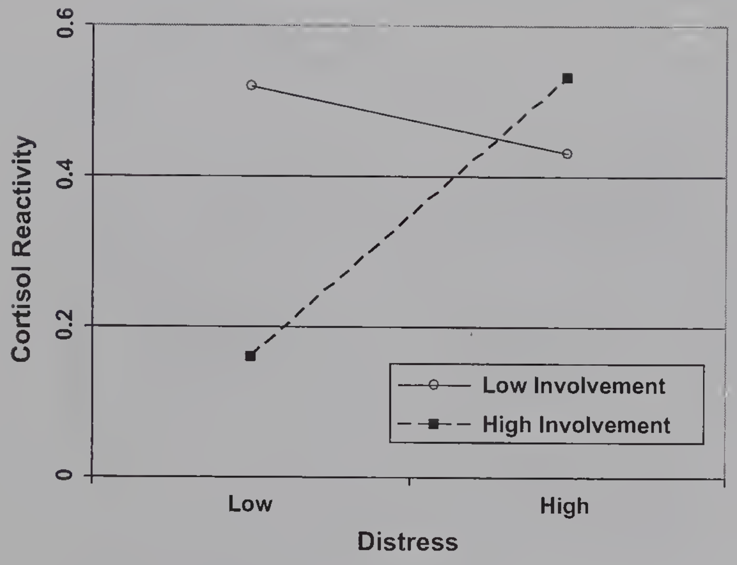


Figure 3. A graphical plot of the significant interaction between observations of children's distress and involvement responses to interparental conflict predicting their cortisol reactivity to the simulated interparental conflict.

potential interplay between children's psychological and cortisol responses to interparental conflict (Davies et al., 2006; Gunnar & Vazquez, 2006; Levine & Weiner, 1988; Spangler & Schieche, 1998). However, little is known about the concurrent interconnections between children's reactivity at these different levels of analysis. Accordingly, this study aimed to break new ground by identifying the nature of the associations between multiple dimensions of children's psychological reactivity to interparental conflict and their cortisol functioning in response to interparental conflict.

Simultaneous analysis of pathways among children's distress, hostile, and involvement responses to interparental conflict and their cortisol reactivity to interparental conflict indicated that distress was significantly and uniquely associated with higher levels of cortisol reactivity to interparental conflict. Confidence in the validity of our findings is bolstered by the considerable stability of pathways between distress responding and cortisol reactivity to conflict across several conditions. First, the specificity of associations between cortisol and distress reactivity to conflict was consistent across measures of child reactivity to conflict that varied in use of method (i.e., observation, questionnaire), informant (i.e., observer, mother), and context (i.e., laboratory, home). Thus, children's heightened cortisol reactivity to simulated interparental conflict was not only specifically associated with observational ratings of children's distress responses within the same conflict context, but it was also predicted by higher levels of maternal questionnaire reports of children's reactivity to interparental conflict in the home. Second, tests of differences in the magnitude of links between maternal reports of the three forms of psychological responding and cortisol reactivity indicated that the association between distress and cortisol reactivity was stronger than comparable paths involving child hostile and involvement responses to conflict. Third, paths between distress and cortisol reactivity were consistent even after including several demographic and child characteristics as confounding factors or moderators in the analyses.

Coherence in the results raises a central question: Why is distress reactivity to conflict uniquely associated with elevated cortisol responding to interparental conflict? Among the conceptualizations of linkages between psychological and physiological responding, the general arousal and negative affect models do not hold the precision necessary to readily explain the associations between distress and cortisol reactivity. In the general arousal model, the assumption that many forms of effortful responding to interparental conflict are coupled with greater LHPA activity would specifically suggest that distress, hostility, and involvement are relatively similar in their roles as predictors of cortisol levels. Correspondingly, the negative affect model does not distinguish between hostility and distress and assumes that negative emotionality is broad correlate of greater cortisol reactivity.

Rather, the distinctive confluence of children's heightened distress and cortisol reactivity to conflict found in the present study supports a distress specificity model and its main premise that distress is distinctly coupled with greater cortisol reactivity (Susman, Dorn, Inoff-Germain, Nottelmann, & Chrousos, 1997). Profiles characterized by elevated distress and LHPA activity in the context of interparental conflict may signify a hyperreactive threat detection system (Gunnar & Vazquez, 2006). Supporting this interpretation, EST posits that elevated distress in response to interparental conflict is a hallmark symptom of children's concerns about security (Cummings & Davies, 1996).

In further informing the distress specificity model, EST also suggests that hypersensitivity to threat may be particularly evident when distress responses to interparental conflict occur in conjunction with high levels of involvement in the conflict. In this context, attempts to intervene in adult problems by children who are distressed by interparental conflict may signify a last resort, defense strategy for preserving security when alternative solutions and resources are not accessible (Davies et al., 2006). Thus, if the blending of distress and involvement does reflect extreme considerable alarm, then distress should be a particularly potent predictor of cortisol reactivity to conflict when child involvement levels are also high. In partially supporting this interpretation, follow-up analyses of an interaction between child distress and involvement revealed that children's distress in the simulated conflict was associated with elevated cortisol reactivity to conflict only when they were also highly involved in the parental problems.

Conceptualizations of threat and challenge help to further address the mechanisms responsible for the confluence of distress and elevated cortisol reactivity. For example, one hypothesis derived from the distress specificity model is that elevations in cortisol are most likely to occur when children cannot access the resources necessary to cope with the stressor or perceive the stressor as beyond their control (Levine & Weiner, 1988; Spangler & Schieche, 1998). Applied to our findings, underlying doubts about the ability to utilize strategies necessary to preserve goals of safety and security may underlie the association between children's distress and adrenocortical reactivity (Saarni, Campos, Camras, & Witherington, 2006). Similarly, social psychological frameworks draw important distinctions between threat and challenge. Whereas the experience of threat is theorized to reflect appraisals that stressful contexts exceed coping resources available to the individual, challenges reflect perceptions that skills and resources available will be sufficient to cope with the demands of the context (Blascovich & Mendes, 2000; Dienstbier, 1989). According to these models, threat is hypothesized to be linked with elevated LHPA activity (Dienstbier, 1989; Mendes, Blascovich, Hunter, Lickel, & Jost, 2007). Integrated with the EST assumption that distress is a key indicator of concerns about safety, one interpretation is that the coupling of distress and cortisol reactivity to conflict may be rooted in the common underlying experience of threat. Although still speculative at this early stage of research, there are some bases for expecting that this response pattern may increase children's adjustment problems, particularly in the form of internalizing symptoms. For example, independent lines of research have identified children's internalizing symptoms as common sequelae of children's heightened adrenocortical reactivity to stressful events (Gunnar & Vazquez, 2006) and their elevated distress responses to interparental conflict (Davies & Cummings, 2006).

Two additional, but inconsistent, findings from our analyses .raise more speculative questions. First, maternal report of child hostility was associated with dampened cortisol reactivity to the simulated interparental conflict. Given that this pathway was only identified in one of the two main analytic models, considerable care should be taken in interpreting the results. Nevertheless, one tentative explanation derived from social psychological conceptualizations of challenge is that high levels of confidence in the ability to weather the stressfulness of interparental conflict may underlie the experience of hostile and dampened cortisol reactions to interparental conflict (e.g., Blascovich & Mendes, 2000). A related interpretation drawn from the attenuation hypothesis is that hostility may be considered dominant, bold, or callous response that may be associated with underarousal of the LHPA system in response to interparental conflict (Susman, 2006). Assuming this finding is replicated in the future, a critical step in future research is to identify the long-term implications of this response pattern. For example, within the attenuation hypothesis (Susman, 2006), the cooccurrence of hostility and dampened cortisol responses may reflect underlying disturbances in neurobiological processing of emotional and fear-relevant parameters of events that ultimately deepen and expand into broader externalizing problems. Second, the findings also indicated that maternal reports of distress were associated with lower preconflict levels of cortisol. However, a comparable path between SPAT distress and preconflict cortisol was not significant. Thus, the findings, as a whole, suggest that distress responses to interparental conflict are not consistent correlates of a broad, trait-like condition of hypercortisolism across multiple contexts. Rather, distress responses to conflict evidenced specificity in its power to predict greater cortisol reactivity to conflict.

Fully interpreting the results of our study also requires consideration of the methodological limitations. Future research on children's coping with family conflict and their cortisol functioning would benefit by increasing the rigor of psychological and physiological assessments. For example, our identification of pathways between distress reactivity and the intercept and slope parameters of cortisol functioning does not afford a thorough analysis of specificity or breadth of the adrenocortical correlates of distress responses to conflict. One important step will be to determine whether distress responding to interparental conflict is specifically associated with cortisol reactivity to interparental conflict or broader indices of cortisol reactivity to a wide array of family, interpersonal, or even more global (i.e., physical and psychosocial) stressors. Likewise, guided by conceptualizations of emotional reactivity in EST, our assessments of emotional distress were largely comprised of assessments of anxiety, fear, and general distress. As a result, our measurement battery did not permit a full analysis of associations between adrenocortical reactivity and other forms of negative emotion such as sadness. In addition, contextualizing the study of cortisol within broader profiles of functioning across multiple physiological systems may also advance psychophysiological models of coping with family conflict (Doussard-Roosevelt, Montgomery, & Porges, 2003; Gordis, Granger, Susman, & Trickett, 2006; Granger et al., 2006). For example, the modest to moderate magnitude of associations between psychological and physiological reactivity to conflict may be augmented by analyses of synchrony in functioning between the LHPA axis and the parasympathetic (e.g., Doussard-Roosevelt et al., 2003) or sympathetic (e.g., Gordis et al., 2006) nervous systems.

Future advances in identifying links between children's behavioral and physiological reactivity to conflict will also benefit from modifications in the temporal spacing of assessments. For example, the number and timing of our cortisol assessments do not permit a definitive analysis of the pattern of cortisol reactivity across stressor and recovery periods (Matthews, Gump, & Owens, 2001). Greater numbers of cortisol assessments in future research would afford flexibility in analyzing both linear and curvilinear trajectories of cortisol over time. Moreover, delineation of individual differences in recovery of cortisol may be achieved by lengthening the temporal spacing between the stressor and some of the cortisol assessments. However, relative to more common methods of assessing cortisol reactivity (e.g., difference scores between preconflict and a single postconflict measure of cortisol), our LGC analysis generated a richer characterization of cortisol reactivity through its ability to incorporate multiple assessments of cortisol that are sensitive to the direction and shape of the trajectory. Likewise, although our use of cross-sectional design is well suited to address the goal of examining the concurrent correspondence between cortisol and behavioral reactivity to conflict, use of longitudinal designs will afford tests of complementary questions about possible transactional influences among behavioral and adrenocortical reactivity over longer time periods.

Finally, although the demographic characteristics of our sample of families were highly comparable to the Northeast and Midwest regions of the United States, the findings in this study were derived from a community sample of predominantly White families. Therefore, the results of our study may not necessarily generalize to families with other racial, ethnic, or high-risk backgrounds. Nevertheless, the failure of a wide array of demographic characteristics to alter as covariates or moderators, the pathways found between children's distress and cortisol reactivity to interparental conflict provides some bases for the generalizability of the results.

In conclusion, our multimethod, multiple level of analysis study is the first foray into identifying the adrenocortical underpinnings of children's behavioral and psychological responses to interparental conflict. Consistent with the distress specificity model in the psychophysiological literature, children's distress responding to interparental conflict was consistently linked with LGC analyses of greater cortisol reactivity to interparental conflict across multiple methods (i.e., observations, questionnaires), informants (i.e., observer, mother), and contexts (i.e., laboratory, home). Moreover, the pathways were distinctive when also taking into consideration other forms of children's reactivity (i.e., hostility, involvement) to conflict and a range of sociodemographic variables as potential confounding factors. In broader accord with EST, children's security in the interparental relationship may be a central goal system organizing the confluence of distress and heightened cortisol reactivity to conflict.

References

Achenbach, T. M. (2006). As others see us: Clinical and research implications of cross-informant correlations for psychopathology. *Current Directions in Psychological Science, 15*, 94–98.

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. London: Sage.

Ashman, S. B., Dawson, G., Panagiotides, H., Yamada, E., & Wilkinson, C. W. (2002). Stress hormone levels of children of depressed mothers. *Development and Psychopathology, 14*, 333–349.

Blascovich, J., & Mendes, W. B. (2000). Challenge and threat appraisals: The role of affective cues. In J. Forgas (Ed.), *Feeling and thinking: The role of affect in social cognition* (pp. 59–82). Paris: Cambridge University Press.

Cahill, L., & McGaugh, J. L. (1998). Mechanisms of emotional arousal and lasting declarative memory. *Trends in Neurosciences, 21*, 294–299.

Cicchetti, D., & Tucker, D. (1994). Development and self-regulatory structures of the mind. *Development and Psychopathology, 6*, 533–549.

Cohen, S., Schwartz, J. E., Epel, E., Kirschbaum, C., Sidney, S., & Seeman, T. (2006). Socioeconomic status, race, and diurnal cortisol decline in the Coronary Artery Risk Development in Young Adults (CARDIA) study. *Psychosomatic Medicine, 68*, 41–50.

Cummings, E. M., & Davies, P. T. (1994). *Children and marital conflict: The impact of family dispute and resolution*. New York: Guilford.

Cummings, E. M., & Davies, P. (1996). Emotional security as a regulatory process in normal development and the development of psychopathology. *Development and Psychopathology, 8*, 123–139.

Cummings, E. M., Davies, P., & Campbell, S. B. (2000). *Developmental psychopathology and family process: Theory, research, and clinical implications*. New York: Guilford.

Cummings, E. M., Schermerhorn, A. C., Davies, P. T., Goeke-Morey, M. C., & Cummings, J. S. (2006). Interparental discord and child adjustment: Prospective investigations of emotional security as an explanatory mechanism. *Child Development*, *77*, 132–152.

Davies, P. T., & Cummings, E. M. (1994). Marital conflict and child adjustment: An emotional security hypothesis. *Psychological Bulletin*, *116*, 387–411.

Davies, P. T., & Cummings, E. M. (2006). Interparental discord, family process, and developmental psychopathology. In D. Cicchetti & D. J. Cohen (Eds.), *Developmental psychopathology: Vol. 3. Risk, disorder, and adaptation* (2nd ed., pp. 86–128). New York: Wiley.

Davies, P. T., Cummings, E. M., & Winter, M. A. (2004). Pathways between profiles of family functioning, child security in the interparental subsystem, and child psychological problems. *Development and Psychopathology*, *16*, 525–550.

Davies, P. T., Forman, E. M., Rasi, J. A., & Stevens, K. I. (2002). Assessing children's emotional security in the interparental subsystem: The security in the interparental subsystem (SIS) scales. *Child Development*, *73*, 544–562.

Davies, P. T., Harold, G. T., Goeke-Morey, M., & Cummings, E. M. (2002). Child emotional security and interparental conflict. *Monographs of the Society for Research in Child Development*, *67*(3, Serial No. 270).

Davies, P. T., Winter, M. A., & Cicchetti, D. (2006). The implications of emotional security theory for understanding and treating childhood psychopathology. *Development and Psychopathology*, *18*, 707–735.

De Los Reyes, A., & Kazdin, A. E. (2005). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin*, *131*, 483–509.

Dienstbier, R. A. (1989). Arousal and physiological toughness: Implications for mental and physical health. *Psychological Review*, *96*, 84–100.

Dettling, A. C., Gunnar, M. R., & Donzella, B. (1999). Cortisol levels of young children in full-day childcare centers: Relations with age and temperament. *Psychoneuroendocrinology*, *24*, 519–536.

Dettling, A. C., Parker, S. W., Lane, S., Sebanc, A., & Gunnar, M. R. (2000). Quality of care and temperament determine changes in cortisol concentrations over the day for young children in child care. *Psychoneuroendocrinology*, *25*, 819–836.

Dickerson, S. S., & Kemeny, M. E. (2004). Acute stressors and cortisol responses: A theoretical integration and synthesis of laboratory research. *Psychological Bulletin*, *130*, 355–391.

Doussard-Roosevelt, J. A., Montgomery, L. A., & Porges, S. W. (2003). Short-term stability of physiological measures in kindergarten children: Respiratory sinus arrhythmia, heart period, and cortisol. *Developmental Psychobiology*, *43*, 231–242.

Fox, N. A., Hane, A. A., & Perez-Edgar, K. (2006). Psychophysiological methods for the study of developmental psychopathology. In D. Cicchetti & D. J. Cohen (Eds.), *Developmental psychopathology: Vol. 2. Developmental neuroscience* (2nd ed., pp. 381–426). New York: Wiley.

Gordis, E. B., Granger, D. A., Susman, E. J., & Trickett, P. K. (2006). Asymmetry between salivary cortisol and alpha - amylase reactivity to stress: Relation to aggressive behavior in adolescents. *Psychoneuroendocrinology*, *31*, 976–987.

Granger, D. A., Cicchetti, D., Rogosch, F. A., Hibel, L. C., Teisl, M., & Flores, E. (2007). Blood contamination in children's saliva: Prevalence, stability, and impact on the measurement of salivary cortisol, testosterone, and dehydroepiandrosterone. *Psychoneuroendocrinology*, *32*, 724–733.

Granger, D. A., Kivlighan, K. T., Blair, C., El-Sheikh, M., Mize, J., Lisonbee, J. A., et al. (2006). Integrating the measurement of salivary alpha -amylase into studies of child health, development, and social relationships. *Journal of Social and Personal Relationships*, *23*, 267–290.

Granger, D. A., Weisz, J. R., & Kauneckis, D. (1994). Neuroendocrine reactivity, internalizing behavior problems, and control-related cognitions in clinic-referred children and adolescents. *Journal of Abnormal Psychology*, *103*, 267–276.

Grych, J. H., & Fincham, F. D. (1990). Marital conflict and children's adjustment: A cognitive-contextual framework. *Psychological Bulletin*, *108*, 267–290.

Gunnar, M. R., & Donzella, B. (2000). Social regulation of the cortisol levels in early human development. *Psychoneuroendocrinology*, *27*, 199–220.

Gunnar, M., Larson, M., Hertsgaard, L., Harris, M., & Brodersen, L. (1992). The stressfulness of separation among 9-month-old infants: Effects of social context variables and infant temperament. *Child Development*, *63*, 290–303.

Gunnar, M. R., & Vazquez, D. M. (2006). Stress neurobiology and developmental psychopathology. In D. Cicchetti & D. J. Cohen (Eds.), *Developmental psychopathology: Vol. 2. Developmental neuroscience* (2nd ed., pp. 533–577). New York: Wiley.

Hertsgaard, L., Gunnar, M., Erickson, M. F., & Nachmias, M. (1995). Adrenocortical responses to the Strange Situation in infants with disorganized/disoriented attachment relationships. *Child Development*, *66*, 1100–1106.

Hibel, L. C., Granger, D. A., Cicchetti, D., & Rogosch, F. (2007). Salivary biomarker levels and diurnal variation: Associations with medications prescribed to control children's problem behaviors. *Child Development*, *78*, 927–937.

Hibel, L. C., Granger, D. A., Kivlighan, K. T., Blair, C., & the Family Life Investigators (2006). Individual differences in salivary cortisol: Associations with common over-the-counter and prescription medication status in infants and their mothers. *Hormones and Behavior*, *50*, 293–300.

Hruschka, D. J., Kohrt, B. A., & Worthman, C. M. (2005). Estimating between- and within-individual variation in cortisol levels using multilevel models. *Psychoneuroendocrinology, 30*, 698–714.

Kagan, J., Reznick, J. S., & Snidman, N. (1998). Biological bases of childhood shyness. *Science, 240*, 167–171.

Katz, L. F. (2001). Physiological processes and mediators of the impact of marital conflict on children. In J. H. Grych & F. D. Fincham (Eds.), *Interparental conflict and child development: Theory, research, and applications* (pp. 188–212). New York: Cambridge University Press.

Kivlighan, K. T., Granger, D., Schwartz, E. B., Nelson, V., Curran, M., & Shirtcliff , E. A. (2004). Quantifying blood leakage into the oral mucosa and its effects on the measurement of cortisol, dehydroepiandrosterone, and testosterone in saliva. *Hormones and Behavior , 46*, 39–46.

Knutson, U., Dahlgreen, J., Marcus, C., Rosberg, S., Bronnegard, M., Stierna, P., et al. (1997). Circadian cortisol rhythms in healthy boys and girls: Relationships with age, growth, body composition, and pubertal development. *Journal of Clinical Endocrinology and Metabolism, 82*, 536–540.

Levine, S., & Weiner, S. G. (1988). Psychoendocrine aspects of mother-infant relationships in nonhuman primates. *Psychoneuroendocrinology, 13*, 143–154.

Loney, B. R., Butler, M. A., Lima, E. N., Counts, C. A., & Eckel, L. A. (2006). The relation between salivary cortisol, callous-unemotional traits, and conduct problems in an adolescent non-referred sample. *Journal of Child Psychology and Psychiatry, 47*, 30–36.

Masharani, U., Shiboski, S., Eisner, M. D., Katz, P. P., Janson, S. L., Granger, D. A., et al. (2005). Impact of exogenous glucocorticoid use on salivary cortisol measurements among adults with asthma and rhinitis. *Psychoneuroendocrinology, 30*, 744–752.

Matthews, K. A., Gump, B. B., & Owens, J. F. (2001). Chronic stress influences cardiovascular and neuroendocrine responses during acute stress and recovery, especially for men. *Health Psychology, 20*, 403–410.

Mendes, W. B., Blascovich, J., Hunter, S. B., Lickel, B., & Jost, J. T. (2007). Threatened by the unexpected: Physiological responses during social interactions with expectancy-violating partners. *Journal of Personality and Social Psychology, 92*, 698–716.

National Institutes of Health. (2003). *Program announcement: Research on children exposed to violence (PAR-03-096)*. Retrieved December 23, 2007, from http://grants.nih.gov/grants/guide/search_results.htm?year=active&scope=pa

Powers, S. I., Pietromonaco, P. R., Gunlicks, M., & Sayer, A. (2006). Dating couples' attachment styles and patterns of cortisol reactivity and recovery in response to a relationship conflict. *Journal of Personality and Social Psychology, 90*, 613–628.

Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interaction effects in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics, 31*, 437–448.

Saarni, C., Campos, J. J., Camras, L. A., & Witherington, D. (2006). Emotional development: Action, communication, and understanding. In N. Eisenberg, W. Damon, & R. M. Lerner (Eds.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (6th ed., pp. 226–299). New York: Wiley.

Salimetrics. (2000). *HS cortisol kit information*. Unpublished manuscript, State College, PA.

Saltzman, K. M., Holden, G. W., & Holahan, C. J. (2005). The psychobiology of children exposed to marital violence. *Journal of Clinical Child and Adolescent Psychology, 34*, 129–139.

Schwartz, E. P., Granger, D. A., Susman, E. J., Gunnar, M. R., & Laird, B. (1998). Assessing salivary cortisol in studies of child development. *Child Development, 69*, 1503–1513.

Spangler, G., & Grossmann, K. E. (1993). Biobehavioral organization in securely and insecurely attached infants. *Child Development, 64*, 1439–1450.

Spangler, G., & Schieche, M. (1998). Emotional and adrenocortical responses of infants to the Strange Situation: The differential function of emotional expression. *International Journal of Behavioral Development, 22*, 681–706.

Stansbury, K., & Gunnar, M. R. (1994). Adrenocortical activity and emotion regulation. In N. A. Fox (Ed.), The development of emotion regulation: Biological and behavioral considerations. *Monographs of the Society for Research in Child Development, 59*(2–3), 108–134.

Susman, E. J. (2006). Psychobiology of persistent antisocial behavior: Stress, early vulnerabilities, and the attenuation hypothesis. *Neuroscience and Biobehavioral Reviews, 30*, 376–389.

Susman, E. J., Dorn, L. D., Inoff-Germain, G., Nottelmann, E. D., & Chrousos, G. P. (1997). Cortisol reactivity, distress behavior, and behavioral and psychological problems in young adolescents: A longitudinal perspective. *Journal of Research on Adolescence, 7*, 81–105.

Tout, K., de Haan, M., Campbell, E. K., & Gunnar, M. R. (1998). Social behavior correlates of cortisol activity in child care: Gender differences and time-of-day effects. *Child Development, 69*, 1247–1262.

Child Development, November/December 2008, Volume 79, Number 6, Pages 1707–1720

Stress Regulation in Adolescents: Physiological Reactivity During the Adult Attachment Interview and Conflict Interaction

Mariëlle D. Beijersbergen, Marian J. Bakermans-Kranenburg, Marinus H. van IJzendoorn, and Femmie Juffer
Leiden University

The current study examined whether adolescents' attachment representations were associated with differences in emotion regulation during the Adult Attachment Interview (AAI; C. George, N. Kaplan, & M. Main, 1996) and during a mother–adolescent conflict interaction task (Family Interaction Task [FIT]; J. P. Allen et al., 2003). Participants were one hundred and fifty-six 14-year-old adolescents. Dismissing adolescents showed less interbeat interval (IBI) reactivity (indicating less stress) during the AAI than secure adolescents. However, during the FIT, dismissing adolescents showed more IBI reactivity. No differences in physiological reactivity were found between individuals with resolved or unresolved loss or trauma during the AAI or FIT. The results indicate that dismissing adolescents may effectively use a defensive strategy during the AAI but less so in direct conflict interaction with their attachment figure.

According to attachment theory, internal working models of attachment influence emotion regulation, both in (early) childhood and in adolescence and adulthood (Cassidy, 1994; Main, Kaplan, & Cassidy, 1985). Physiological parameters provide an excellent opportunity to test this hypothesized link (Spangler & Zimmermann, 1999), but research in this area is still scarce. Using physiological measures, the current study examined whether adolescents' working models of attachment (or attachment representations) are associated with their emotion regulation during two situations in which the attachment system is activated: during the Adult Attachment Interview (AAI; George, Kaplan, & Main, 1996; Hesse, 1999; Main, Goldwyn, & Hesse, 2003) and during a dyadic (mother–adolescent) conflict interaction task (Allen et al., 2003; Kobak, Sudler, & Gamble, 1991; Strodtbeck, 1951).

Emotion regulation patterns can be seen as part of an adaptive strategy with the goal of maintaining the relationship with the attachment figure (Cassidy, 1994). Individuals are suggested to have flexible or inflexible styles of emotion regulation, developed as the result of particular caregiving histories. Infants with secure attachment relationships as well as adults with secure attachment representations are characterized by open, flexible emotional expressions (Bretherton, 1990; Cassidy, 1994). Individuals with insecure attachment representations typically show a restricted range of emotions. Dismissing persons are suggested to systematically suppress emotions; they would mask negative affect. Evidence for the nature of this defensive strategy is still limited. Nevertheless, the first studies using physiological measures (Dozier & Kobak, 1992; Roisman, Tsai, & Chiang, 2004) suggest that dismissing individuals experience stress, although it is not displayed overtly. Preoccupied individuals, in contrast, maximize the expression of negative emotions (Cassidy, 1994; Kobak, Holland, Ferenz-Gillies, Fleming, & Gamble, 1993; Main, 1990). Even though these emotion regulation strategies may be adaptive in the relationship with the attachment figure (not being rejected or gaining attention from an unavailable caregiver, respectively), they may be maladaptive in other contexts and have negative psychological and developmental consequences (Cassidy, 1994; Cassidy & Kobak, 1988; Main, 1990).

Internal working models of attachment may also have an impact on the regulation of attention. Dismissing individuals are hypothesized to systematically exclude attachment-relevant information from consciousness (see Bowlby, 1980; Main, 1999). As a consequence, dismissing persons would usually be unable to give evidence for what they claim was a perfectly normal or very nice childhood. Preoccupied persons, on the contrary, show a strong focus on

---

Mariëlle D. Beijersbergen is now working at the Radboud University Nijmegen Medical Centre.

Support from the Netherlands Organization for Scientific Research to the second author (NWO VIDI grant) and to the third author (NWO SPINOZA Prize) is gratefully acknowledged. We also acknowledge the financial support received from Wereldkinderen and Stichting Kind en Toekomst to F.J. We thank the adolescents and their parents for their participation. Our thanks are also to Nicole Bimmel and Karin Dofferhoff who contributed to various phases of the project.

Correspondence concerning this article should be addressed to Marian J. Bakermans-Kranenburg, Centre for Child and Family Studies, Leiden University, P.O. Box 9555, 2300 RB Leiden, The Netherlands. Electronic mail may be sent to bakermans@fsw.leidenuniv.nl.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7906-0012

attachment relationships and experiences (Hesse, 1999). Moreover, it has been suggested that during information processing, their attention is centered on negative emotions (Spangler & Zimmermann, 1999).

In contrast to individuals with organized attachment classifications, unresolved individuals show a momentary breakdown in their strategy (Hesse, 1999; Lyons-Ruth & Jacobvitz, 1999; Main et al., 2003). Adults are classified as unresolved when they show lapses in the monitoring of reasoning or discourse or report extreme behavioral reactions when talking about loss or other trauma (Main et al., 2003). It has been proposed that these adults may be impaired in emotion regulation when confronted with traumatic experiences (Spangler & Zimmermann, 1999).

Emotions affect physiological responses (Brownley, Hurwitz, & Schneiderman, 2004; Dawson, Schell, & Filion, 2004; Hagemann, Waldstein, & Thayer, 2003). Physiological responses are a result of the activity of the autonomic nervous system (ANS; Porges, 1995). This system consists of two subsystems: the parasympathetic nervous system (PNS) and the sympathetic nervous system (SNS). Both originate in the brainstem and influence the regulation of organs such as heart, lungs, and kidneys, as well as sweat glands, and blood vessels. The PNS is involved in growth and restorative processes in the body. The SNS quickly mobilizes existing reserves of the body when a situation requires a fight-or-flight reaction. To get insight into the activity of the ANS in situations in which emotions are elicited, measures like interbeat interval (IBI), heart rate variability (root mean of the squared successive differences [RMSSD]), and skin conductance level (SCL) have been used (Bradley, 2004).

IBI is an indicator for the time between two consecutive beats of the heart. When a person is under stressful behavior challenge (or, e.g., exercising), his or her heart rate may quicken; therefore, IBI will be shorter. RMSSD may be lower under the same challenge (Brownley et al., 2004). The level of electrodermal activity (EDA) as indexed by SCL is influenced by increases and decreases in sweat in the eccrine sweat glands (Boucsein, 1992; Dawson et al., 2004). More emotional arousal will result in higher SCL levels (Bradley, 2004). These physiological measures are suggested to be indices for stress (see Boucsein, 1992, for a review on EDA as indicator of stress; Porges, 1995) but may also be indicative of emotions like happiness or sadness, or indicative of attention, depending on the context in which these measures are used (Bradley, 2004). In adequate (stressful) contexts, they can provide a window on (negative) emotions that may or may not be expressed overtly. It is important to note that IBI, RMSSD, and SCL are differentially influenced by the SNS and PNS: IBI is influenced by both (Brownley et al., 2004), whereas RMSSD is primarily influenced by the PNS (Hagemann et al., 2003), and SCL is influenced by the SNS (Dawson et al., 2004). These measures are not interchangeable; they may show different patterns when used in the same context (e.g., Roisman et al., 2004).

In attachment research, the potential of psychophysiological parameters is increasingly used for enhancing our understanding of emotion regulation in individuals with different attachment patterns (e.g., Oosterman & Schuengel, 2007; Spangler & Grossmann, 1993; Sroufe & Waters, 1977; Zelenko et al., 2005). Only two studies, however, have examined physiological regulation during the AAI. Dozier and Kobak (1992) hypothesized that deactivation (a strategy preferred by individuals with a more dismissing representation), as contrasted with hyperactivation (more characteristic of preoccupied representations) would be related to SCL reactivity during the AAI because of its relation to behavioral inhibition (Fowles, 1980). Their findings suggested that individuals using deactivating strategies were indeed more stressed during the AAI. Roisman et al. (2004) extended this line of research by including cardiovascular reactivity (i.e., IBI reactivity) as well as skin conductance reactivity during the AAI. Convergent with their expectations, deactivation was related to SCL reactivity but not to cardiovascular reactivity (which is suggested to be indicative of behavioral activation; see Fowles, 1980). In conclusion, both studies showed an association between deactivation and increased stress during the AAI as indicated by sympathetic reactivity.

Differences in emotion regulation patterns may not only be evident during the AAI, they may also appear in other situations in which the attachment system is activated, such as in dyadic conflict interactions (Allen et al., 2003; Feeney & Cassidy, 2003). During discussions of disagreements with their parents, adolescents need to establish autonomy while also maintaining relatedness. In such conflict situations, secure individuals are supposed to be better able to balance exploration (e.g., autonomy strivings) and attachment (e.g., relatedness) than insecure individuals (Allen & Land, 1999). Focusing on physiological reactivity in adults during a discussion with their romantic partners, Roisman (2007) found that deactivation was related to SCL reactivity, whereas hyperactivation was associated with HR reactivity. Attachment security was related to less SCL reactivity during interactions with romantic partners. To our knowledge, no studies have yet investigated the link between attachment and stress reactivity during interactions between parents and their adolescent children.

The relation between attachment representation and patterns of interaction during conflict resolution has been examined before. Allen and Hauser (1996) reported that young adults' coherence of discourse during the AAI could be predicted from their mothers' promoting autonomy and relatedness 11 years earlier. Using Kobak's Q-sort (1993), Allen et al. (2003) revealed that dyadic relatedness shown in conflict interactions between mothers and 16-year-old adolescents was related to adolescent attachment security. Kobak et al. (1993) demonstrated that secure adolescents and their mothers were characterized by less dysfunctional anger and less avoidance of problem solving during conflict interactions. More dysfunctional anger as well as more maternal dominance were displayed in interactions between adolescents with deactivating strategies and their mothers.

In the current study, we examined physiological reactivity during the AAI and during an interaction task in which mothers and adolescents tried to reach consensus in an area of disagreement. It was expected that dismissing adolescents would experience more stress during the AAI than secure participants because of their hypothesized defensive strategy during the task of reflecting on early attachment experiences. In accordance with previous studies, no differences between preoccupied and secure adolescents were expected. Furthermore, unresolved individuals might show more reactivity during the questions concerning loss and (other) trauma. Regarding the conflict interaction task (Family Interaction Task [FIT]), we hypothesized that dismissing as well as preoccupied individuals would be more stressed than adolescents with secure attachment representations. No differences were expected between adolescents with resolved or unresolved states of mind because it seems unlikely that the interaction task on a disagreement would trigger memories of loss or trauma. Finally, it was expected that secure adolescents would show more autonomous relatedness during conflict interaction than insecure adolescents.

Method

Participants

Participants were one hundred and fifty-six 14-year-old internationally adopted adolescents who took part in a longitudinal study that started in infancy (Jaffari-Bimmel, Juffer, van IJzendoorn, Bakermans-Kranenburg, & Mooijaart, 2006; Juffer, Bakermans-Kranenburg, & van IJzendoorn, 2005). We report on 152 participants because in 2 cases, the AAIs of the adolescents could not be coded due to technical problems and 2 other AAIs were not classifiable because the respondents did not understand the questions due to (very) low intellectual level (IQs of 58 and 82, respectively).

Sixty-eight boys and 84 girls were involved in the current study (see Table 1). All children were adopted before the age of 6 months (M = 10.0 weeks, SD = 5.30). They came from Sri Lanka (n = 94), South Korea (n = 38), and Colombia (n = 20). The adoptive families predominantly belonged to middle class or upper middle class (Jaffari-Bimmel et al., 2006). Mean age of the adoptive mothers at the time of the birth of the children was 33.1 (SD = 3.55, N = 142) and of the adoptive fathers 35.0 (SD = 3.55, N = 141).

Procedure

The adoptive families were randomly recruited through Dutch adoption organizations and were followed from infancy to adolescence. The current study reports on the data collected at a follow-up at 14 years of age. Adolescents participated in 3.5-hr home visits together with their mothers, except for four families where the fathers participated. Results were similar when these fathers were excluded from the analyses. During the visits, the AAI was administered to the adolescents, as well as an intelligence test. The adolescents also completed questionnaires and participated in an interaction task with their mothers. One questionnaire, used as baseline for the physiological measures during the AAI, and the AAI were completed in a separate room, without the mother present. A second questionnaire, used as baseline for the physiological measures during the FIT, was completed in the presence of the mother. During the entire session, the adolescents were connected to the VU-AMS recording device (the Vrije Universiteit Ambulatory Monitoring System [AMS 36]; Groot, De Geus, & De Vries, 1998; see also Bimmel, van IJzendoorn, Bakermans-Kranenburg, Juffer, & De Geus, 2008).

Measures

AAI. The AAI (Main et al., 2003) is an hour-long, semistructured interview that assesses an individual's current state of mind with respect to attachment. In this interview, respondents were asked about their childhood experiences with their adoptive parents and how they thought they were affected by them. Other questions concerned experiences of loss and trauma. Finally, respondents were invited to describe possible changes in the relationship with their adoptive parents since childhood and the current

Table 1
Descriptive Statistics: Demographics and Background Variables

	Secure[a] (*n* = 57)		Dismissing[b] (*n* = 62)		Preoccupied[c] (*n* = 33)		Resolved[d] (*n* = 110)		Unresolved[e] (*n* = 23)		Total[f] (*N* = 152)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Health condition at placement	0.36	0.14	0.36	0.14	0.38	0.11	0.36	0.14	0.37	0.13	0.37	0.13
Age at placement (in weeks)	9.9	5.40	10.6	5.45	8.8	4.77	10.1	5.40	9.3	4.57	10.0	5.30
Age at time of assessment	14.5	0.54	14.3	0.48	14.5	0.62	14.4	0.55	14.4	0.51	14.4	0.53
IQ	100.8	12.74	100.4	13.02	100.5	14.18	100.3	12.94	99.4	11.98	100.6	13.09
Socioeconomic status	10.0	2.60	10.1	2.64	10.0	2.82	9.8	2.79	10.5	2.18	10.0	2.65
Smoking	2.1	1.27	2.1	1.47	2.0	1.30	2.1	1.35	2.0	1.22	2.1	1.35
Number of sports	1.4	0.89	1.4	1.10	1.5	1.16	1.4	1.04	1.7	1.06	1.4	1.03
Number of words during AAI[g]	1,907	599.0	1,356	5,29.1	2,823	1,425.7	1,821	998.5	2,261	1,095.3	1,881	992.2
	N	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%
Sex[e] (*N* male)	19	33.3	34	54.8	15	45.5	47	42.7	12	52.2	68	44.7
Country of origin												
Sri Lanka	31	54.4	38	61.3	25	75.8	66	60.0	16	69.6	94	60.9
South Korea	18	31.6	15	24.2	5	15.2	31	28.2	2	8.7	38	25
Colombia	8	14	9	14.5	3	9.1	13	11.8	5	21.7	20	13.2

Note. As we only included adolescents who experienced loss or trauma in the resolved and unresolved categories, the total *n* for this subgroup is 133. AAI = Adult Attachment Interview.
[a] Range *n* = 54–57. [b] Range *n* = 61–62. [c] Range *n* = 30–33. [d] Range *n* = 108–110. [e] Range *n* = 22–23. [f] Range *N* = 147–152. [g] Number of words used during responses to AAI questions 6–11, except for 9a, which is about abuse.

relationship with them. On basis of verbatim transcripts of the AAI, the adolescents were judged as having a secure (F), dismissing (Ds), or preoccupied (E) attachment representation. The discourse of secure individuals is coherent. They are able to freely value their experiences and yet stay objective regardless of the nature of their experiences. Individuals with an insecure attachment representation significantly violate Grice's (1975) criteria for coherence. Dismissing individuals typically idealize their parents and claim lack of memory for their childhood. In rare cases, they derogate their attachment experiences. Individuals with a preoccupied state of mind are still confused and overwhelmed by their childhood experiences. They are angry toward their parents or their discourse is characterized by vague speech (Main et al., 2003). On top of their main classification, individuals may be classified as unresolved-disorganized (U) when they show lapses in the monitoring of reasoning or discourse in reaction to loss or other traumatic events (Main et al., 2003).

Previous studies provided preliminary evidence for the validity of the AAI in adolescent samples. For example, continuity of attachment from infants' Strange Situation classifications (Ainsworth, Blehar, Waters, & Wall, 1978) to adolescents' AAI classifications has been reported (e.g., Hamilton, 2000; Main, Hesse, & Kaplan, 2005). In addition, adolescent attachment representations as assessed with the AAI were associated with several developmental outcomes in the predicted way (e.g., Adam, Sheldon-Keller, & West, 1996; Mayseless & Scharf, 2007). Although these studies usually involved adolescents with a mean age of 16 years or older, one study reported considerable stability of the AAI classifications when 10-year-olds were reassessed 4 years later (Ammaniti, van IJzendoorn, Speranza, & Tambelli, 2000). In addition, using Kobak's Q-sort (1993), Allen and colleagues. (Allen, Porter, McFarland, Boykin-McElhaney, & Marsh, 2007; Allen, Porter, McFarland, Marsh, & Boykin-McElhaney, 2005) found associations between 14-year-olds' attachment representations and mother–adolescent and father–adolescent interactions 1 year earlier. The validity of the AAI in (early) adolescence thus looks promising.

The AAIs were coded by the first author. For interrater reliability, 18 randomly selected interviews were also coded by the second author. Interrater agreement was 78% ($\kappa = .64$) for three-way classifications (secure, dismissing, and preoccupied) and 83% ($\kappa = .77$) for four-way classifications (secure, dismissing, preoccupied, and unresolved). Disagreements between coders were resolved by discussion. Intraclass correlation for the overall unresolved score was $r = .89$. Finally, a continuous dismissing score was derived by using the maximum score of an individual on the following scales: idealization of mother, idealization of father, derogation of mother, derogation of father, and overall derogation of attachment. Intraclass correlation for the dismissing score was $r = .71$.

Of the 152 adopted adolescents, 57 (37.5%) were secure, 62 (40.8%) dismissing, and 33 (21.7%) preoccupied. Taking the unresolved category into account, the distribution was: 50 (32.9%) secure, 57 (37.5%) dismissing, 19 (12.5%) preoccupied, and 26 (17.1%) unresolved. The distribution of classifications in our sample differed significantly from the normative distribution in nonclinical adolescent samples (van IJzendoorn & Bakermans-Kranenburg, 2008) for the three-way distribution, $\chi^2(2, N = 152) = 30.74, p < .01$, as well as for the four-way distribution, $\chi^2(3, N = 152) = 15.36, p < .01$. In the current sample, the insecure categories were overrepresented.

In order to test the validity of the AAI in the present sample, we investigated whether emotional investment in others versus self was related to adolescents' AAI classifications. Because dismissing individuals tend to emphasize their independence and dismiss the importance of attachment relationships (Hesse, 1999), they were expected to emotionally invest less in others than nondismissing individuals who value attachment relationships (secure individuals) or feel emotionally very dependent on others (preoccupied individuals; see Mikulincer & Shaver, 2007). Our findings confirmed this hypothesis (Beijersbergen, van IJzendoorn, Bakermans-Kranenburg, & Juffer, 2007), providing additional support for the validity of the AAI when used with adolescents.

Attachment classification was not associated with gender, country of birth, socioeconomic status (SES), health condition or age at adoptive placement, age at time of assessment, number of sports, smoking, or intelligence. A significant association was found between number of words used during the AAI and attachment classification, with preoccupied individuals using the most words, whereas dismissing individuals gave the shortest answers, $F(2, 149) = 33.85, p < .01$.

FIT. Using a revealed differences task (Strodtbeck, 1951), we investigated the patterns of interaction between adolescents and their mothers (Allen et al., 2003; Kobak et al., 1991). Mothers and adolescents were asked to discuss and try to reach consensus on an issue on which they disagreed during 10 min. The discussion started after they listened to a tape that had been recorded in absence of the mother, on which the adolescent stated his or her opinion about the disagreement, as well as the opinion the adolescent thought the mother had.

The interactions were coded with the autonomy and relatedness coding system of Allen and colleagues (Allen et al., 1994). Adolescents received scores ranging from 0 to 4 on four scales (derived from nine subscales): (a) *exhibiting autonomy*, states reasons clearly for disagreeing, confidence in stating thoughts and opinions; (b) *inhibiting autonomy*, recanting a position, overpersonalizing, pressures to agree; (c) *exhibiting relatedness*, validates/agrees/positively reacts to other person, engaged interaction; and (d) *inhibiting relatedness*, distracting/ignoring, hostile/devaluing statements. The subscale *recanting a position* was excluded from the inhibiting autonomy scale because it was not associated with the other two subscales as a consequence of lack of variance. Because the inhibiting autonomy and inhibiting relatedness scales were strongly correlated ($r = .72$, $p < .01$), they were combined into one scale, *inhibiting autonomy relatedness*. The scales concerning exhibiting autonomy and exhibiting relatedness were not combined because the correlation between these scales was modest ($r = .29, p < .01$). The discussions were coded by two different coders who were unaware of other characteristics of the dyads and who were both trained by an expert who received training from Dr. J. P. Allen. Mean intercoder reliability between the expert and the two coders was .79 (range = .68 – .92, $n = 30$). Internal consistency of the three major scales used in the analyses was adequate (exhibiting autonomy $\alpha = .82$, exhibiting relatedness $\alpha = .60$, and inhibiting autonomy relatedness $\alpha = .82$).

Table 2 presents means and standard deviations of the autonomy-relatedness scales. Scale scores were missing for 1 participant because no parent was present during the session. The autonomy-relatedness scales for one other adolescent could not be scored because of technical problems with the recording.

Physiological measures. The VU-AMS recording device (Groot et al., 1998) was used to measure IBI, RMSSD (an index for heart rate variability), and galvanic SCL. These measures were selected to replicate (IBI and SCL) and extend (RMSSD) prior work by Dozier and Kobak (1992) and Roisman et al. (2004). We assume that IBI reactivity, RMSSD reactivity, and SCL reactivity scores of the adolescents are indices of stress because these measures were used during two potentially stressful tasks.

Table 2
Scores on Autonomy-Relatedness Scales of the Family Interaction Task (FIT) per Attachment Classification

	Secure ($n = 56$)[a] M (SD)	Dismissing ($n = 61$)[a] M (SD)	Preoccupied ($n = 33$) M (SD)	Resolved ($n = 108$)[b] M (SD)	Unresolved ($n = 23$) M (SD)	Total ($N = 150$)[b] M (SD)
Exhibiting autonomy	2.39 (0.69)	2.00 (0.75)[c]	2.35 (0.84)	2.26 (0.78)	2.17 (0.71)	2.22 (0.76)
Exhibiting relatedness	1.65 (0.59)	1.64 (0.60)	1.67 (0.62)	1.61 (0.57)	1.79 (0.66)	1.65 (0.60)
Undermining autonomy relatedness	1.07 (0.78)	0.86 (0.72)	1.03 (0.74)	0.94 (0.73)	1.10 (0.82)	0.98 (0.75)

Note. As we only included adolescents who experienced loss or trauma in the resolved and unresolved categories, the total N for this subgroup is 133.
[a] FIT is missing for one participant in this group. [b] FIT is missing for two participants in this group. [c] *Mean difference dismissing – secure* = 0.39, *SE* = 0.14, $p < .05$.

Three disposable electrocardiogram electrodes were placed on the adolescent's chest: The first was placed at the jugular notch of the sternum, between the collarbones; the second was placed below the left breast, 4 cm (1.5 inch) under the nipple, between two ribs; the last electrode was placed at the right side of the chest, between the lower two ribs. The SCL electrodes were placed on the thenar and hypothenar eminences of the palms of the hands. The VU-AMS device continuously recorded IBI. RMSSD was calculated based on the raw IBI data and was sampled every 10 s. SCL was sampled every 500 ms. The quality of the signal and attachment of the electrodes were checked by online monitoring of the physiological data. The VU-AMS device failed to record physiological data during one home visit. SCL recordings were unreliable for 1 participant, and another participant had unreliable physiological recordings during the AAI. For one adolescent, the equipment failed to record physiology after the introduction of the FIT. The corresponding data were excluded from the analyses.

Physiological values during the answers to the following AAI questions were used in the analyses (George et al., 1996): (6) When you were upset when you were little, what did you do, what would happen? Can you think of specific incidents? Physically hurt? Ill? (7) Could you describe your first separation from your parents? (8) Did you ever feel rejected as a child? What did you do? Do you think your parents realized they were rejecting you? (8a) Were you ever frightened or worried as a child? (9) Were your parents ever threatening with you—maybe for discipline or jokingly? (9a) Some people have memories of some kind of abuse. Did that ever happen to you, or in your family? (10) How do you think your overall experiences have affected your adult personality? (10a) Are there any aspects to your early experiences that you feel were a setback in your development? (11) Why do you think your parents behaved as they did, during your childhood? (13) Did you experience the loss of a parent or other close loved one? (14) Have you had any other experiences which you would regard as potentially traumatic? These questions were selected because we expected that they would show the largest differences in physiological reactivity between the organized secure and insecure attachment strategies (e.g., questions 6–9 and 10–11) or between persons with an resolved or unresolved state of mind (e.g., questions 9a, 13, and 14). The last 3 min of the episode in which the adolescents were alone completing a questionnaire was used as baseline for the AAI. Because 7 adolescents did not fill in the questionnaire, they could not be included in the analyses.

During the FIT, physiological values were recorded for 4.5 min starting from the moment the dyad began the discussion. After 4.5 min, a number of mothers and adolescents drifted away from the original task and began a conversation about nonproblem issues, for example, about the day at school. The last 3 min of the episode in which the adolescents completed a questionnaire in the presence of the mother was used as baseline for the FIT. Two adolescents did not fill in this questionnaire and were therefore excluded from the analyses.

Implausible physiological values were deleted (De Geus, 1996; Groot et al., 1998). Physiological reactivity was calculated per AAI question for IBI, RMSSD, and SCL by subtracting means of the AAI baselines from the means during the selected AAI questions. Physiological reactivity during the FIT was computed by subtracting the FIT baseline from the means during the FIT. More reactivity, and thus more stress, is indicated by higher SCL, lower IBI, and lower RMSSD difference scores. Covariates (as indicated under the AAI subsection) were only included if they were associated with the dependent as well as the independent variables. Following Keppel and Wickens's (2004) recommendation concerning extreme scores, we included outliers in the analyses. Results remained similar when outliers were changed into the next most extreme scores (Tabachnick & Fidell, 2001).

Intelligence. Intelligence was included because it may affect heart rate: Higher IQ has been associated with heart rate deceleration (Lewis & Wilson, 1970; but see Farrington, 1997, for an exception). The adolescents completed the abbreviated Groningen Intelligence Test (Luteijn & Van der Ploeg, 1983). The following three subsets were included: cipher, enumerate words, and word matrices. Mean IQ score was 100.6 (*SD* = 13.09).

Physical condition and smoking. Physical health and smoking influence heart rate (De Geus, Boomsma, & Snieder, 2003; Farrington, 1997; Vander, Sherman, & Luciano, 2001). We therefore asked participants in which sports they had been active in the previous year. Participants were also asked to rate on a 5-point scale how much they smoked (1 = *never* to 5 = *often*).

Health condition at placement. Health condition at adoptive placement was used as an index for the health condition of the infant from birth to placement in the family (Stams, Juffer, & van IJzendoorn, 2002). Information for this index was gathered in the first interview with the parents, which was conducted when the infants were 5 months old. Health condition at placement was calculated by the standardized summation of (a) birth weight, (b) incidence of prematurity, and (c) health problems at placement (reversely coded). Higher scores represent better health condition at placement.

Socioeconomic status. SES of the adoptive families was assessed when the children were 7 years old, combining the educational and vocational background of both parents (for more details, see Stams et al., 2002). Scores for SES correspond to socioeconomic strata as follows: 3–9 lower class, 9–12 middle class, and 12–16 upper class. Adoptive families had a mean SES of 10.0 (*SD* = 2.65).

Data Analysis

First, analyses were conducted for physiological reactivity during the AAI. Linear mixed models were run with attachment representation, question, and the interaction between question and attachment representation as fixed effects. We also tested whether subject should be included as a random effect and which covariance structure should be used for the error term. For each physiological measure, the best fitting models were selected using Akaike's information criterion and log likelihood (Fitzmaurice, Laird, & Ware, 2004). Post hoc tests were conducted to test whether secure individuals significantly differed from dismissing and preoccupied individuals with regard to physiological reactivity during the AAI. Finally, we calculated correlations between the dismissing score and physiological reactivity.

For the analyses concerning unresolved loss, we selected the participants who did report loss or trauma during the AAI (n = 133). Because all adolescents were adopted before the age of 6 months and they were asked about their own memories concerning loss or trauma, all reported experiences concerned postadoption loss or trauma. Linear mixed model analyses were conducted with the classification unresolved with respect to loss or trauma, question, and the interaction between question and unresolved attachment as fixed effects. Correlations were calculated between the overall unresolved score (for either loss or other trauma) and physiological reactivity.

Next, physiological reactivity during the FIT was examined. We conducted the same analyses (three-way classifications and resolved vs. unresolved classifications; dismissing score and unresolved score) for the AAI data. Finally, reactivity during the selected AAI questions was averaged to compare reactivity during AAI and FIT. We tested whether one task caused more physiological reactivity than the other, whether persons with different attachment classifications differed in reactivity during both tasks, and whether there was an interaction effect between task and attachment.

Results

Preliminary Analyses

With a multivariate analysis of variance, we tested for significant differences in autonomy relatedness between adolescents with divergent attachment classifications. The overall effect was not significant, $F(6, 290) = 1.60, p = .15$. However, because we had a priori hypotheses regarding the different types of adolescents' interactive behaviors and the sphericity assumption held, more powerful focused univariate analyses were also conducted (Keppel & Wickens, 2004). A significant effect was found for exhibiting autonomy, $F(2, 147) = 4.56, p < .05$. Secure adolescents ($EM = 2.39$, $SE = 0.10$) had higher scores on this scale than dismissing adolescents ($EM = 2.00, SE = 0.10$). Exhibiting autonomy during the FIT was also negatively associated with the AAI dismissing scale ($r = -.18, p < .05$). No association was found between the dismissing scale and showing relatedness ($r = -.03, p = .69$) or inhibiting autonomy relatedness ($r = -.03, p = .68$).

Unresolved and not-unresolved adolescents showed no differences in autonomy-relatedness scores, $F(3, 127) = 2.42$, $p = .07$. Moreover, none of the correlations between the autonomy-relatedness scales and the unresolved score in the group of adolescents who experienced loss or trauma ($n = 131$) was significant (exhibiting autonomy $r = .06$, $p = .49$; exhibiting relatedness $r = .17$, $p = .06$; inhibiting autonomy relatedness $r = .16$, $p = .08$).

Physiological Reactivity During the AAI

Means and standard deviations of the raw physiological data during baseline as well as during the AAI questions are presented in Table 3. Neither secure, dismissing, and preoccupied participants nor resolved and unresolved participants had significantly different baselines for IBI, RMSSD, or SCL.

Associations between physiological reactivity during the AAI and background variables. A significant relation was found between IBI reactivity and age at time of the assessment. Furthermore, RMSSD reactivity was related to smoking and gender. SCL reactivity was associated with the following variables: age at adoptive placement, age at time of assessment, country of origin, and number of words used during the AAI. A table presenting all statistics concerning the relations between physiological reactivity and background variables for the total group as well as for the group who experienced loss or trauma is available upon request.

Table 3

Physiological Assessments: Raw Values for the Baseline and AAI Questions

	IBI[a]		RMSSD[b]		SCL[c]	
Question	*M*	*SD*	*M*	*SD*	*M*	*SD*
Baseline	779.85	92.73	54.33	26.72	11.26	7.96
Upset	761.31	94.37	48.76	23.54	11.54	7.95
Separation	761.68	92.78	49.75	24.18	11.52	7.65
Rejection	779.64	99.11	50.69	25.77	11.39	7.44
Frightened	764.80	96.55	50.69	26.44	11.58	7.69
Threatened	776.75	101.40	50.74	26.60	11.62	7.55
Effects	770.32	95.10	50.19	23.93	11.73	7.50
Setback	780.90	105.11	52.98	28.35	11.54	7.47
Why behaved	769.39	96.14	51.24	26.41	11.65	7.49
Abuse	786.45	105.33	52.14	28.26	11.66	7.45
Loss	770.83	98.49	52.39	27.58	11.67	6.97
Other trauma	774.44	98.04	52.82	28.77	11.79	7.26

Note. IBI = interbeat interval; RMSSD = root mean of the squared successive differences; SCL = skin conductance level.
[a] $N = 150$, except for baseline IBI ($N = 141$). [b] N ranges between 147 and 149, except for baseline RMSSD ($N = 139$). [c] $N = 149$, except for baseline SCL ($N = 142$).

For the adolescents who experienced loss or other trauma, we found a relation between IBI reactivity and intelligence. In this subgroup, SCL reactivity was associated with age at time of assessment, number of sports, and number of words used during the AAI. Because number of words used during the AAI was related to SCL reactivity as well as to the three-way attachment classification, we used this variable as a covariate in all pertinent analyses.

Differences between secure, dismissing, and preoccupied adolescents. A linear mixed model for IBI reactivity with a diagonal covariance structure and subject included as a random effect showed that attachment representation, question, and the interaction between question and attachment representation were all significant, $F(2, 140) = 3.37$, $p < .05$; $F(7, 235) = 14.09$, $p < .01$; and $F(14, 235) = 1.77$, $p < .05$, respectively. Adolescents appeared to be most reactive during the questions related to being upset and being separated (respectively, $M = 17.86$, $SD = 51.64$; $M = 18.12$, $SD = 50.79$). In contrast, they showed least reactivity during the setback question ($M = -2.17$, $SD = 53.97$). Post hoc tests revealed that dismissing participants showed less IBI reactivity than secure ones (M *difference* $= -22.60$, $SE = 8.91$, $df = 140$, $p < .05$), indicating that they were less stressed than secure adolescents (see Figure 1). No differences were found between secure and preoccupied adolescents (M *difference* $= -6.29$, $SE = 10.84$, $df = 140$, $p = .99$). Finally, the interaction between attachment and

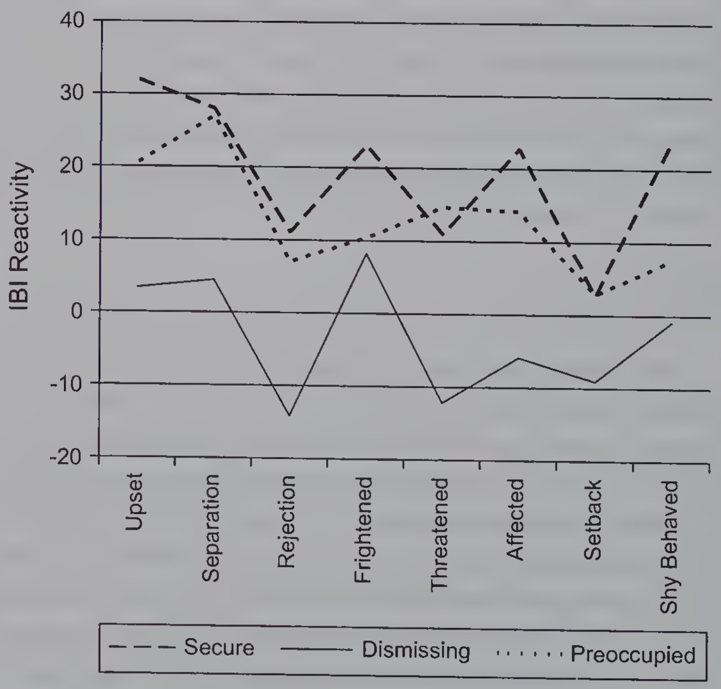


Figure 1. Interbeat interval reactivity of secure, dismissing, and preoccupied adolescents during the Adult Attachment Interview.

question revealed that for the questions on separation, $t(201) = 2.27$, $p < .05$, and threat, $t(218) = 2.24$, $p < .05$, preoccupied adolescents showed an increase in IBI reactivity, whereas secure individuals displayed a decrease in IBI reactivity.

Adolescents with secure, dismissing, or preoccupied classifications did not differ significantly from each other on RMSSD reactivity, $F(2, 137) = 0.81$, $p = .45$, or on SCL reactivity, $F(2, 141) = 1.35$, $p = .26$. Participants displayed more SCL reactivity during the question how they were affected by their childhood than during the question on rejection (*M difference* = 0.36, *SE* = 0.11, *df* = 585, $p < .05$). When the outliers were changed into the next most extreme scores, the adolescents also showed more SCL reactivity during the frightening and threatening questions than during the rejection question (*M difference* = 0.23, *SE* = 0.07, *df* = 912, $p < .05$; *M difference* = 0.27, *SE* = 0.09, *df* = 905, $p < .05$). Because the results of the analyses with SCL reactivity did not change whether or not number of words used during the AAI was included, only the statistics of the analyses without number of words as covariate are presented.

Dismissing score. Controlling for background variables that were associated with the physiological measures and the dismissing score, we found that higher scores on the dismissing scale were associated with lower IBI reactivity during the questions concerning being upset and when asked for effects of childhood experiences ($r = -.18$, $p < .05$; $r = -.22$, $p < .05$, respectively). A trend was found for the relation between the dismissing score and the question about separations ($r = -.16$, $p = .06$). When the outliers were changed into the next most extreme scores, the correlation between the dismissing scale and the IBI reactivity during the question about rejection was also significant ($r = -.17$, $p < .05$). Adolescents with higher dismissing scores tended to show less reactivity during these questions. RMSSD and SCL reactivity were not related to the dismissing scale during any of the selected AAI questions (RMSSD separate scores for gender: range $r = -.23$ to .19, $p = ns$; SCL range $r = -.04$ to $-.08$, $p = ns$).

Do adolescents with and without unresolved attachment classifications differ in physiological reactivity during the AAI? The results for SCL reactivity remained the same when number of words was included in the analysis as a covariate, we therefore only report statistics of the linear mixed model for SCL reactivity without number of words. The linear mixed models for IBI, RMSSD, and SCL had a diagonal covariance structure and included subject as a random effect. Question was a significant predictor for IBI, $F(2, 158) = 9.69$, $p < .001$: All adolescents showed more IBI reactivity ($M = 12.45$, $SD = 49.09$) during the loss question, indicating that this question was more stressful than the questions concerning abuse ($M = 6.32$, $SD = 53.69$) and other trauma ($M = 9.90$, $SD = 53.39$). In none of the tests, unresolved attachment with respect to loss or trauma was a significant predictor of physiological reactivity, IBI, $F(1, 127) = 0.42$, $p = .52$; RMSSD, $F(1, 120) = 0.42$, $p = .52$; and SCL, $F(1, 123) = 0.03$, $p = .86$. Physiological reactivity was neither related to the unresolved loss score.

Physiological Reactivity During the FIT

In order to keep the analyses concise and focused, we decided to follow through only on IBI reactivity during the FIT because adolescents with different attachment representations differed in IBI reactivity during the AAI. Mean IBI baseline value was 794.11 ($SD = 110.46$), whereas mean IBI value during the FIT was 775.63 ($SD = 104.82$).

Secure, dismissing, and preoccupied adolescents. Means of IBI reactivity for the three attachment classifications were 9.43 ($SD = 46.81$) for secure adolescents, 30.90 ($SD = 58.33$) for dismissing adolescents, and 9.87 ($SD = 52.00$) for preoccupied adolescents. The contrast for secure versus dismissing adolescents showed a significant difference (difference = 21.48, $SE = 9.84$, $p < .05$): Dismissing adolescents showed more reactivity, which suggests they felt more stressed than secure adolescents during the FIT. The dismissing scale was significantly associated with IBI reactivity ($r = .18$, $p < .05$).

IBI reactivity during the FIT and unresolved loss/trauma. No significant differences were found for IBI reactivity between adolescents with or without an unresolved attachment classification, $F(1, 127) = 0.87$, $p = .35$, and the unresolved score was not correlated to IBI reactivity during the FIT ($r = -.03$, $p = .76$, $n = 129$).

IBI Reactivity During the AAI and the FIT

A linear mixed model was run with IBI reactivity as dependent variable and task and attachment classification as independent variables. There was no significant main effect for task, $F(1, 280) = 1.33$, $p = .25$, nor for attachment, $F(2, 280) = .07$, $p = .93$. The interaction between task and attachment was significant, $F(2, 280) = 6.18$, $p < .01$. As can be seen in Figure 2, in comparison with the secure adolescents, dismissing adolescents showed more IBI reactivity during the FIT, whereas they displayed less reactivity during the AAI, $t(280) = -3.34$, $p < .01$.

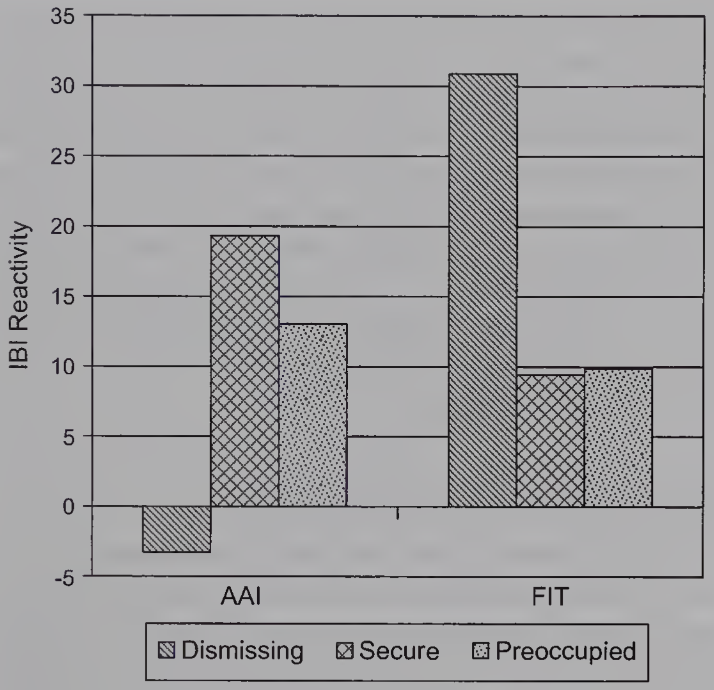


Figure 2. Differences in interbeat interval reactivity during the Adult Attachment Interview (AAI) and during the Family Interaction Task. (FIT).

Discussion

The current study investigated whether attachment classification was related to physiological reactivity during the AAI (Main et al., 2003) and during a dyadic interaction task (FIT; Allen et al., 2003; Kobak et al., 1991; Strodtbeck, 1951). IBI reactivity suggested that dismissing participants were less stressed during the AAI than secure adolescents. However, during the conflict interaction task, involving a discussion with their mother on an issue on which they disagreed, dismissing adolescents seemed to be more stressed. Furthermore, during the AAI questions about separation and threat, preoccupied individuals showed elevated IBI reactivity, whereas secure individuals displayed decreases in reactivity. We found no differences in physiological reactivity between adolescents with a resolved or unresolved state of mind during the AAI or during the FIT. In addition, there were no differences in IBI reactivity between the AAI and the FIT, and there was no overall effect for attachment. However, a significant interaction effect indicated that in comparison with secure adolescents, dismissing adolescents were more stressed during the FIT and less stressed during the AAI.

Our study extends previous research with a broader spectrum of physiological measures during the AAI. Moreover, we examined not only associations with secure, dismissing, and preoccupied attachment classifications but also tested whether individuals with and without unresolved loss showed different physiological reactivity during the pertinent AAI questions on loss and (other) trauma. We used Main et al.'s (2003) classification system instead of Kobak's Q-sort (1993) as was done in previous studies. Furthermore, the present study was conducted in the home; whereas prior studies conducted the AAI in a laboratory setting, that setting may be more stressful than the home. In addition, according to our knowledge, no other studies have yet focused on physiological reactivity shown by persons with different attachment representations during a mother–adolescent conflict interaction task. Finally, our sample consisted of adolescents. The present findings thus extend previous physiological investigations. Nevertheless, further studies are necessary to draw more definite conclusions regarding physiological reactivity in attachment-relevant situations, in adolescents as well as in adults.

Dismissing adolescents did not seem to experience more stress during the AAI, whereas they did appear to be more stressed than secure adolescents during the FIT as indicated by IBI reactivity. Both tasks place different demands on the participants. During the AAI, participants are asked to produce their childhood memories and reflect on them (Hesse, 1999). Our findings suggest that dismissing adolescents are less open to the challenge of the AAI than secure adolescents and are able to cope with the interview in a somewhat superficial manner. They might therefore display less IBI reactivity during the AAI, but it seems impossible to be uninvolved and detached during a direct interaction task with their mother with the goal of reaching consensus in an area of disagreement. We thus propose that the defensive strategy of dismissing adolescents might be effective during the AAI even at an early stage of information processing, and that they are less open to seriously address these intimate questions compared to secure adolescents. This hypothesis has been suggested before. In a study using the Stroop test (Zeijlmans Van Emmichoven, van IJzendoorn, De Ruiter, & Brosschot, 2003), persons with a secure attachment representation showed slower response latencies than insecure persons. Moreover, clinical subjects with a secure attachment representation had the largest response times to threat words. The authors therefore hypothesized that secure individuals may be more open to this type of threatening information that is carefully processed, whereas dismissing individuals exclude this unbalancing information at an early stage.

It should be noted that this hypothesis of excluding attachment-related memories and experiences at an early stage of information processing may pertain to adolescents only, as prior work on adults showed a different picture. Adolescents are different from adults in that they did not have had much time to work through their attachment experiences and are in the process of becoming less dependent on their parents (Allen & Land, 1999; van IJzendoorn & Bakermans-Kranenburg, 2008). Adolescents also differ from adults in that their brain is not yet fully developed (see for a review, Blakemore & Choudhury, 2006). Maturation of the frontal cortex continues into adolescence. Magnetic resonance imaging studies have shown changes in white and gray matter during adolescence, which may be indicative of increased axonal myelination and synapse proliferation. These changes may account for the lower capacity of adolescents to control and coordinate their thoughts and behavior (executive function, including selective attention, decision making, voluntary response inhibition, and working memory) as compared to adults. Adolescents may thus show a different pattern of associations between AAI representations and (physiological) responses because of their transitional life stage (becoming less dependent on their parents) and their less developed cognitive functioning (less mature frontal cortex).

In contrast to their lower physiological reactivity during the AAI, dismissing adolescents appeared to be more stressed than secure adolescents during the FIT. When the dismissing adolescents are in direct interaction with their mother, they may have less opportunity to effectively use defensive processes, and they may thus experience more stress than secure adolescents. In the FIT, the stakes may be felt to be higher because the discussion pertains to present real-life conflict issues, and the mother may play a more demanding or provocative role than an unknown interviewer who is not able to check the validity of any response. This result converges with a study of Roisman (2007) in which deactivation was associated with SCL reactivity during the discussion of a disagreement with the participant's romantic partner.

Differences between dismissing and secure adolescents during the FIT may also be due to their mothers' contribution during this task. Mothers' interactions during the discussion vary across dyads and may be dependent on the adolescent's behavior. The interactive behavior of mothers of dismissing adolescents might be more stress evoking than that of mothers of their secure counterparts. In fact, during the FIT, both dismissing adolescents and their potentially dismissing or insensitive mothers contribute to the conflict interaction which might exacerbate the tensions in the relationship compared to stress felt during the unilateral AAI narrative.

Preoccupied adolescents differed from adolescents with a secure attachment representation only in their response to the questions on separation and threat during the AAI. During the FIT they did not differ in reactivity from secure adolescents. The strategy of these individuals is to maximize attention to attachment relationships and experiences (Hesse, 1999). The AAI and the FIT thus seem not to challenge preoccupied adolescents like they do dismissing individuals, which fits nicely with the idea that preoccupied persons are used to talk about relationships and emotional experiences. In addition, the FIT provides preoccupied individuals with the undivided attention of their attachment figure. During the AAI questions regarding separation and threat, these individuals may be especially triggered to recall anxious experiences and as a consequence show elevated IBI reactivity. Roisman (2007) showed that hyperactivation was associated with more heart rate reactivity during a discussion between romantic partners. However, relationships between adolescents and their parents and between romantic partners have different characteristics. Romantic partners may decide to end the relationship which raises intense feelings of anxiety in preoccupied individuals, whereas even preoccupied adolescents may always feel the strong bond with their parents—whether or not this bond is (in)secure.

We found differences in cardiac reactivity as opposed to SCL reactivity. Because no difference in RMSSD reactivity (which is an indicator of parasympathetic activation) was found between adolescents with a dismissing or secure attachment representation, we tentatively speculate that differences in IBI reactivity may be mainly due to differences in sympathetic activation (see Roisman, 2007). Activation of the sympathetic branch has been associated with deactivation of attachment before (e.g., Dozier & Kobak, 1992; Roisman, 2007; Roisman et al., 2004). This seems to be in contrast with the result that SCL reactivity was not associated with adolescent attachment; however, the lack of findings for SCL may be (partly) a consequence of the various ethnicities represented in the current sample. SCL is suggested to be influenced by a person's ethnicity; for example, White participants tend to have higher SCL levels than Black participants (Boucsein, 1992). As our sample consisted of adolescents with different ethnic backgrounds, this may have been a problem for our SCL recordings. Although we controlled for possible associations between SCL and country of origin, a more specific measure for ethnicity may be needed as even adolescents from the same country may vary widely in skin color.

We found no differences in physiological reactivity between resolved and unresolved adolescents. This might be due to the way we measured physiological reactivity. We only focused on reactivity during the loss, abuse, and trauma questions but experiences of loss and trauma may also be discussed in other parts of the AAI. Moreover, the breakdown in strategy by which unresolved loss is characterized is usually very brief (Hesse, 1999; Hesse & Main, 2000), consequently physiological changes may also have been momentary rather than during the entire response to these questions. Future research should try and connect the moment of breakdown in speech during the AAI with the recordings of physiological reactivity.

Although the adolescents in our sample were adopted, they may not be too different from other adolescents. They were adopted in infancy, at a very early age (before 6 months), and were not selected on special needs. Their mean IQ score was not different from the norm for 14- to 15-year-olds, $t(150) = 0.57$, $p = .57$. Moreover, although they had less optimal scores for inhibiting autonomy relatedness compared to a high school sample (Allen & Hauser, 1996), they exhibited more optimal autonomy behaviors compared to an academic low-risk group (Boykin-McElhaney & Allen, 2001). The adoptive status of the adolescents may, however, (partly) account for the relatively high percentage of insecure attachment representations in our sample. The adopted adolescents experienced the loss of their birth parents and possibly other attachment figures, even when they do not explicitly remember them. In addition, their adoptive status may be an issue in particular during adolescence because of the process of identity formation in this stage of life (Brodzinsky, Schechter, & Henig 1992; but see Juffer & van IJzendoorn, 2007).

A limitation of the current study may be that during the baseline periods, adolescents completed a questionnaire, whereas they answered interview questions or were involved in a discussion during the two target tasks. Speaking versus completing a questionnaire may differentially influence physiological activity (e.g., Berntson et al., 1997). However, we were not so much interested in the comparison of physiological response during baseline and these two tests. Our focus pertained to physiological differences in reactivity between adolescents with different attachment representations. We controlled for number of words when necessary, but the findings remained the same.

In sum, the current study is the first to investigate physiological reactivity in adopted adolescents during the AAI as well as during a dyadic conflict interaction task. We propose that dismissing adolescents appear to be able to effectively use their defensive strategy during the AAI as they seem to be less stressed than adolescents with a secure attachment representation, but they appear to experience more stress than secure adolescents in direct interaction with their mothers around a conflict issue. Attachment representations thus play an important role in emotion regulation in attachment-relevant conflict situations, also during adolescence.

References

Adam, K. S., Sheldon-Keller, A. E., & West, M. (1996). Attachment organization and history of suicidal behavior in clinical adolescents. *Journal of Consulting and Clinical Psychology, 64*, 264–272.

Ainsworth, M. D. S., Blehar, M., Waters, E., & Wall, S. (1978). *Patterns of attachment: A psychological study of the Strange Situation.* Hillsdale, NJ: Erlbaum.

Allen, J. P., & Hauser, S. T. (1996). Autonomy and relatedness in adolescent-family interactions as predictors of young adults' states of mind regarding attachment. *Development and Psychopathology, 8*, 793–809.

Allen, J. P., Hauser, S. T., Bell, K. T., McElhaney, K. B., Tate, D. C., Insabella, G. M., et al. (1994). *Autonomy and relatedness coding system manual, version 2.14.* Unpublished manuscript, University of Virginia.

Allen, J. P., & Land, D. (1999). Attachment in adolescence. In J. Cassidy & P. R. Shaver (Eds.), *Handbook of attachment: Theory, research, and clinical applications* (pp. 319–335). New York: Guilford.

Allen, J. P., McElhaney, K. B., Land, D. J., Kuperminc, G. P., Moore, C. W., O'Beirne-Kelly, H., et al. (2003). A secure base in adolescence: Markers of attachment security in the mother-adolescent relationship. *Child Development, 74*, 292–307.

Allen, J. P., Porter, M., McFarland, C., Boykin-McElhaney, K., & Marsh, P. (2007). The relation of attachment security to adolescents' paternal and peer relationships, depression, and externalizing behavior. *Child Development, 78*, 1222–1239.

Allen, J. P., Porter, M. R., McFarland, F. C., Marsh, P., & Boykin-McElhaney, K. (2005). The two faces of adolescents' success with peers: Adolescent popularity, social adaptation, and deviant behavior. *Child Development, 76*, 747–760.

Ammaniti, M., van IJzendoorn, M. H., Speranza, A. M., & Tambelli, R. (2000). Internal working models of attachment during late childhood and early adolescence: An exploration of stability and change. *Attachment and Human Development, 2*, 328–346.

Beijersbergen, M. D., Van IJzendoorn, M. H., Bakermans-Kranenburg, M. J., & Juffer, F. (2007). *Validity of the Adult Attachment Interview in adolescents: Associations with conflict interactions, emotional investment, and relational support.* Unpublished manuscript.

Berntson, G. G., Bigger, J. T., Eckberg, D. L., Grossmann, P., Kaufmann, P. G., Malik, M., et al. (1997). Heart rate variability: Origins, methods, and interpretive caveats. *Psychophysiology, 34*, 623–648.

Bimmel, N., van IJzendoorn, M. H., Bakermans-Kranenburg, M. J., Juffer, F., & De Geus, E. J. C. (2008). Problem behavior and heart rate reactivity in adopted adolescents: Longitudinal and concurrent relations. *Journal of Research on Adolescence, 18*, 201–214.

Blakemore, S.-J., & Choudhury, S. (2006). Development of the adolescent brain: Implications for executive function and social cognition. *Journal of Child Psychology and Psychiatry, 47*, 296–312.

Boucsein, W. (1992). *Electrodermal activity.* New York: Plenum.

Bowlby, J. (1980). *Attachment and loss: Vol. 3. Loss: Sadness and depression.* New York: Basic Books.

Boykin-McElhaney, K., & Allen, J. P. (2001). Autonomy and adolescent social functioning: The moderating effect of risk. *Child Development, 72*, 220–235.

Bradley, M. M. (2004). Emotion and motivation. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (2nd ed., pp. 602–642). Cambridge, UK: Cambridge University Press.

Bretherton, I. (1990). Open communication and internal working models: Their role in the development of attachment relationships. In R. A. Thompson (Ed.), *Nebraska symposium on motivation: Socioemotional development* (pp. 57–113). Lincoln: University of Nebraska Press.

Brodzinsky, D. M., Schechter, M. D., & Henig, R. M. (1992). *Being adopted: The lifelong search for self.* New York: Anchor Books.

Brownley, K. A., Hurwitz, B. E., & Schneiderman, N. (2004). Cardiovascular psychophysiology. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (2nd ed., pp. 224–264). Cambridge, UK: Cambridge University Press.

Cassidy, J. (1994). Emotion regulation: Influences of attachment relationships. *Monographs of the Society for Research in Child Development, 59*(2-3, Serial No. 240), 228–249.

Cassidy, J., & Kobak, R. R. (1988). Avoidance and its relation to other defensive processes. In J. Belsky & T. Nezworski (Eds.), *Clinical implications of attachment* (pp. 300–323). Hillsdale, NJ: Erlbaum.

Dawson, M. E., Schell, A. M., & Filion, D. L. (2004). The electrodermal system. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (2nd ed., pp. 200–223). Cambridge, UK: Cambridge University Press.

De Geus, E. J. C. (1996). *Processing lbl-files created with amsgra: Reading the ambulatory HR, RMSSD/HRV and motility data.* Retrieved May 17, 2006, from http://www.psy.vu.nl/vu-ams/example/index.html

De Geus, E. J. C., Boomsma, D. I., & Snieder, H. (2003). Genetic correlation of exercize with heart rate and respiratory sinus arrhythmia. *Official Journal of the American College of Sports Medicine, 35*, 1287–1295.

Dozier, M., & Kobak, R. R. (1992). Psychophysiology in attachment interviews: Converging evidence for deactivating strategies. *Child Development, 63*, 1473–1480.

Farrington, D. P. (1997). The relationship between low resting heart rate and violence. In A. Raine, P. Brennan, D. P. Farrington, & S. A. Mednick (Eds.), *Biosocial bases of violence* (pp. 89–106). New York: Plenum.

Feeney, B. C., & Cassidy, J. (2003). Reconstructive memory related to adolescent-parent conflict interactions: The influence of attachment-related representations on immediate perceptions and changes in perceptions over time. *Journal of Personality and Social Psychology, 85*, 945–955.

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2004). *Applied longitudinal analysis.* New York: Wiley.

Fowles, D. C. (1980). The three arousal model: Implications of gray's two-factor learning theory for heart rate, electrodermal activity, and psychopathy. *Psychophysiology, 17*, 87–104.

George, C., Kaplan, N., & Main, M. (1996). *Adult Attachment Interview* (3rd ed.). Unpublished manuscript, University of California at Berkeley.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics III: Speech acts* (pp. 41–58). New York: Academic Press.

Groot, P. F. C., De Geus, E. J. C., & De Vries, J. (1998). *Ambulatory monitory system. User manual v1.2* (AMS device 36xx). Amsterdam, The Netherlands: Vrije Universiteit, Department of Psychophysiology.

Hagemann, D., Waldstein, S. R., & Thayer, J. F. (2003). Central and autonomic nervous system integration in emotion. *Brain and Cognition, 52*, 79–87.

Hamilton, C. E. (2000). Continuity and discontinuity of attachment from infancy through adolescence. *Child Development, 71*, 690–694.

Hesse, E. (1999). The Adult Attachment Interview: Historical and current perspectives. In J. Cassidy & P. R. Shaver (Eds.), *Handbook of attachment: Theory, research, and clinical applications* (pp. 395–433). New York: Guilford.

Hesse, E., & Main, M. (2000). Disorganization in infant and adult attachment: Descriptions, correlates, and implications for developmental psychopathology. *Journal of the American Psychoanalytic Association, 48*, 1097–1127.

Jaffari-Bimmel, N., Juffer, F., van IJzendoorn, M. H., Bakermans-Kranenburg, M. H., & Mooijaart, A. (2006). Social development from infancy to adolescence: Longitudinal and concurrent factors in an adoption sample. *Developmental Psychology, 42*, 1143–1153.

Juffer, F., Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2005). The importance of parenting in the development of the disorganized attachment: Evidence from a preventive intervention study in adoptive families. *Journal of Child Psychology and Psychiatry, 46*, 263–274.

Juffer, F., & van IJzendoorn, M. H. (2007). Adoptees do not lack self-esteem: A meta-analysis of studies on self-esteem of transracial, international, and domestic adoptees. *Psychological Bulletin, 133*, 1067–1083.

Keppel, G., & Wickens, T. D. (2004). *Design and analysis: A researcher's handbook* (4th ed.). Upper Saddle River, NJ: Pearson Education.

Kobak, R. R. (1993). *The Attachment Q-Sort.* Unpublished manuscript, University of Delaware.

Kobak, R. R., Holland, E. C., Ferenz-Gillies, R., Fleming, W. S., & Gamble, W. (1993). Attachment and emotion regulation during mother-teen problem solving: A control theory analysis. *Child Development, 64,* 231–245.

Kobak, R. R., Sudler, N., & Gamble, W. (1991). Attachment and depressive symptoms during adolescence: A developmental pathway analysis. *Development and Psychopathology, 3,* 461–474.

Lewis, M., & Wilson, C. D. (1970). The cardiac response to a perceptual cognitive task in the young child. *Psychophysiology, 6,* 411–420.

Luteijn, F., & Van der Ploeg, F. A. E. (1983). *Groninger Intelligentie Test.* Lisse, Netherlands: Swets & Zeitlinger.

Lyons-Ruth, K., & Jacobvitz, D. (1999). Attachment disorganization: Unresolved loss, relational violence, and lapses in behavioral and attentional strategies. In J. Cassidy & P. R. Shaver (Eds.), *Handbook of attachment: Theory, research, and clinical applications* (pp. 520–554). New York. Guilford.

Main, M. (1990). Cross-cultural studies of attachment organization: Recent studies, changing methodologies, and the concept of conditional strategies. *Human Development, 33,* 48–61.

Main, M. (1999). Epilogue. Attachment theory: Eighteen points with suggestions for future studies. In J. Cassidy & P. R. Shaver (Eds.), *Handbook of attachment: Theory, research, and clinical applications* (pp. 845–887). New York: Guilford.

Main, M., Goldwyn, R., & Hesse, E. (2003). *Adult attachment scoring and classification systems.* Unpublished manuscript, University of California at Berkeley.

Main, M., Hesse, E., & Kaplan, N. (2005). Predictability of attachment behavior and representational processes at 1, 6, and 19 years of age: The Berkeley longitudinal study. In K. E. Grossmann, K. Grossmann, & E. Waters (Eds.), *Attachment from infancy to adulthood: The major longitudinal studies* (pp. 245–304). New York: Guilford.

Main, M., Kaplan, N., & Cassidy, J. (1985). Security in infancy, childhood and adulthood: A move to the level of representation. In I. Bretherton & E. Waters (Eds.), Growing points in attachment theory and research. *Monographs of the Society for Research in Child Development, 50*(1-2, Serial No. 209), 66–106.

Mayseless, O., & Scharf, M. (2007). Adolescents' attachment representations and their capacity for intimacy in close relationships. *Journal of Research on Adolescence, 17,* 23–50.

Mikulincer, M., & Shaver, P. R. (2007). *Attachment patterns in adulthood: Structure, dynamics, and change.* New York: Guilford.

Oosterman, M., & Schuengel, C. (2007). Physiological effects of separation and reunion in relation to attachment and temperament in young children. *Developmental Psychobiology, 49,* 119–128.

Porges, S. W. (1995). Cardiac vagal tone—A physiological index of stress. *Neuroscience and Biobehavioral Reviews, 19,* 225–233.

Roisman, G. I. (2007). The psychophysiology of adult attachment relationships: Autonomic reactivity in marital and premarital interactions. *Developmental Psychology, 43,* 39–53.

Roisman, G. I., Tsai, J. L., & Chiang, K. S. (2004). The emotional integration of childhood experiences: Physiological, facial expressive, and self-reported emotional response during the Adult Attachment Interview. *Developmental Psychology, 40,* 776–789.

Spangler, G., & Grossmann, K. E. (1993). Biobehavioral organization in securely and insecurely attached infants. *Child Development, 64,* 1439–1450.

Spangler, G., & Zimmermann, P. (1999). Attachment representation and emotional regulation in adolescents: A psychobiological perspective on internal working models. *Attachment and Human Development, 1,* 270–290.

Sroufe, L. A., & Waters, E. (1977). Heart rate as a convergent measure in clinical and developmental research. *Merill-Palmer Quarterly, 23,* 3–27.

Stams, G. J. M., Juffer, F., & van IJzendoorn, M. H. (2002). Maternal sensitivity, infant attachment, and temperament in early childhood predict adjustment in middle childhood: The case of adopted children and their biologically unrelated parents. *Developmental Psychology, 38,* 806–821.

Strodtbeck, F. L. (1951). Husband-wife interaction over revealed differences. *American Sociological Review, 16,* 468–473.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Boston: Allyn & Bacon.

Vander, A., Sherman, J., & Luciano, D. (2001). *Human physiology: The mechanisms of body function.* New York: McGraw-Hill.

van IJzendoorn, M. H., & Bakermans-Kranenburg, M. J. (2008). The distribution of adult attachment representations in clinical groups: Meta-analytic search for patterns of attachment. In H. Steele & M. Steele (Eds.), *Clinical applications of the Adult Attachment Interview* (pp. 69–98). New York: Guilford.

Zeijlmans Van Emmichoven, I. A., van IJzendoorn, M. H., De Ruiter, C., & Brosschot, J. F. (2003). Selective processing of threatening information: Effects of attachment representation and anxiety disorder on attention and memory. *Development and Psychopathology, 15,* 219–237.

Zelenko, M., Kraemer, H., Huffman, L., Gschwendt, M., Pageler, N., & Steiner, H. (2005). Heart rate correlates of attachment status in young mothers and their infants. *Journal of the American Academy of Child and Adolescent Psychiatry, 44,* 470–476.

Child Development, November/December 2008, Volume 79, Number 6, Pages 1721 – 1735

Effects of Aggression on Achievement: Does Conflict With the Teacher Make It Worse?

Deborah Stipek and Sarah Miles
Stanford University

This longitudinal study examined different explanations for negative associations between aggression and academic achievement using data collected from 403 children from low-income families followed from kindergarten or first grade (ages 6 and 7 years) through fifth grade (ages 10 – 11 years). Most results of growth curve analyses examining change over time and path analyses examining associations among the variables within grades were consistent with the hypothesis that the effect of aggression on achievement was partially mediated by the conflictual relationships relatively more aggressive children tended to develop with their teachers and concomitant reductions in engagement in academic tasks. The evidence suggested, however, that the relationship between aggression and achievement is complex and reciprocal. Gender differences were also observed.

Countless studies have found associations between children's social behavior and their academic performance (for an illustration of a study that did not find these associations, see Duncan et al., 2007). Less well understood is *how* social behavior affects students' learning. This study was designed to examine evidence for different explanations for the association, including the role that children's relationships with their teachers plays in mediating the effects of social behavior on students' academic engagement and achievement.

We focused on aggressive behavior, a form of externalizing against peers. Recent studies on children's aggression have distinguished between overt physical aggression, found more commonly among boys than girls, and "relational aggression" (social or indirect aggression, such as excluding a child from play), found more commonly among girls (e.g., Crick, Grotpeter, & Bigbee, 2002; see Crick et al., 1999, for a review). The current study focuses on what is usually referred to as physical aggression. Because we relied on teacher ratings, we did not differentiate aggression based on different reasons—instrumental versus reactive—as a few researchers have done in past studies (e.g., Little, Brauner, Jones, Nock, & Hawley, 2003).

In a review focusing on externalizing behavior, Hinshaw (1992) summarizes evidence on four possible explanations for associations with achievement: (a) achievement affects externalizing behavior, (b) externalizing behavior affects achievement, (c) there is a reciprocal relationship between externalizing behavior and achievement, and (d) some third variable affects both externalizing behavior and achievement similarly. Analyses for the current study are based on this framework, extended to assess mediators of the association between aggression and achievement.

Evidence

Looking across extant studies, there is evidence for all of Hinshaw's (1992) explanations. First, results of some studies suggest that achievement affects students' social behavior. Miles and Stipek (2006), for example, found that poor literacy achievement in first and third grades predicted relatively high aggressive behavior in third and fifth grades, respectively. In a study on Australian children, Jorm, Share, Matthews, and Maclean (1986) found no differences in children's antisocial behavior in kindergarten as a function of reading skills, but the children who had reading difficulties were significantly more antisocial than the normal readers by the end of first and second grades. Williams and McGee (1994) similarly found that boys with reading disabilities at age 9 were more likely to develop a conduct disorder at age 15.

The data came from the school transition study, which followed approximately 400 low-income children and families from kindergarten or the first grade through the fifth grade. Principal investigators include Deborah Stipek, Penny Hauser-Cram, Walter Secada, Heather Weiss, and Jennifer Greene. Funding for the study was received from the MacArthur Foundation, the Department of Education, and the Foundation for Child Development. The authors are grateful to Tricia Valeski for her contributions to the conceptualization of this study.

Correspondence concerning this article should be addressed to Deborah Stipek, School of Education, Stanford University, 485 Lasuen Mall, Stanford, CA 94305-3096. Electronic mail may be sent to stipek@stanford.edu.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7906-0013

Finally, McGee, Williams, Share, Anderson, and Silva (1986) reported that children who were having considerable difficulty learning to read at the beginning of school already showed behavior problems. Behavior problems also increased from ages 5 to 9 at a much faster pace for children who had serious reading difficulties than for other children.

The present study investigates further evidence related to possible effects of achievement on aggression, employing growth curve analysis to assess the degree to which change in achievement from kindergarten or first grade to fifth grade on an experimenter-implemented test predicted change in the amount of aggression rated by teachers over those grades.

There is also evidence for Hinshaw's (1992) second explanation that conduct problems affect academic achievement. Aggression and other externalizing behaviors may directly affect engagement and learning. For example, children who are aggressive may spend relatively more time misbehaving or being disciplined, reducing the amount of time they have to spend engaged in academic work. Consistent with this proposal, Coie and Dodge (1988) found that first- and third-grade students who were aggressive were more likely to be reprimanded by the teacher and spent less time "on task" than other children. Arnold (1997) similarly observed that misbehavior (aggression, hostility, and noncompliance) predicted low on-task behavior in their study of 4- to 6-year-old boys. The misbehavior prompted teachers to remove boys from the activity, and presumably as a consequence, they spent less time learning the material presented during that lesson. Accordingly, the present study assessed the degree to which changes in aggression predicted changes in achievement.

Aggressive behavior may also undermine engagement and learning indirectly. In this study, we test the hypothesis that the effect of aggression on students' engagement and learning is mediated by the relationships students develop with their teachers. Specifically, we propose that aggressive children form conflictual relationships with teachers, which presumably diminish their desire to engage in the work the teacher gives them and thus undermines their academic engagement and achievement. Our hypothesis is based on a set of empirical findings, summarized in the following, which support each of the links in the model we propose. Taken together, the evidence suggests the model, but it has not been previously tested directly.

Our prediction that children who are relatively more aggressive would develop more negative relationships with their teachers is consistent with previous research showing that teachers claim to have more conflictual relationships with children who exhibit antisocial behaviors (e.g., Birch & Ladd, 1998; Hamre & Pianta, 2001; Howes, Phillipson, & Peisner-Feinberg, 2000; Ladd, Birch, & Buhs, 1999; Ladd & Burgess, 2001; Pianta & Steinberg, 1992; Silver, Measelle, Armstrong, & Essex, 2005), and one study in which low-income kindergarten children's own ratings of the amount of conflict they had with their teachers were associated with teachers' ratings of their conduct problems (Mantzicopoulos, 2005).

The causal relationship between children's social behavior and their relationships with teachers is most likely reciprocal. For example, Hamre and Pianta (2001) found that among children who were identified as having behavior problems in kindergarten, those who developed relationships with teachers characterized by low levels of conflict and dependency had fewer discipline problems and more positive motivation than their counterparts through eighth grade. Silver et al. (2005) reported that conflict in the teacher–child relationship in kindergarten predicted faster rates of increase in aggressive behavior from kindergarten through third grade. To examine the possibility of a reciprocal relationship, we assess how well changes in teacher–child conflict predicted changes in aggressive behavior, as well as how well changes in aggressive behavior predicted teacher–child conflict.

Several studies have demonstrated associations between teacher–child relationships and academic performance for elementary-aged children (Birch & Ladd, 1997; Furrer & Skinner, 2003; Ladd & Burgess, 2001; Mantzicopoulos, 2005; Pianta & Stuhlman, 2004) and adolescents (e.g., Crosnoe, Johnson, & Elder, 2004; Gregory & Weinstein, 2004). Connell and his colleagues (Connell, 1990; Connell & Wellborn, 1991; Skinner & Belmont, 1993; see also Ryan & Deci, 2000) developed a self-system model to explain the effect of teacher–child relationships on children's learning behavior and achievement. According to Connell, children's motivation to engage in particular activities is determined in part by whether these activities take place in a context that satisfies their need for belonging or being socially connected. Children who feel a sense of belonging and social support adopt or internalize the goals and values of their teachers, such as active engagement and participation in school work (Deci & Ryan, 1985; Wentzel, 1997; see also National Research Council Committee on Increasing High School Students' Engagement and Motivation to Learn, 2004). In contrast, children who have conflict with their teachers are likely to feel alienated and consequently disengage from sanctioned classroom behavior (Connell, 1990; Lynch & Cicchetti, 1997).

There is some empirical evidence to support the proposed effects of negative relationships with teachers on adaptive participation (e.g., Birch & Ladd, 1998; Ladd et al., 1999). For example, in a study following children from kindergarten through eighth grade, Hamre and Pianta (2001) found a negative association between teacher – child conflict in kindergarten and children's positive work habits in first through fourth grades and with disciplinary problems in fifth and sixth grades. Ladd and Burgess (2001) found that children who had "chronic" levels of teacher – child conflict in kindergarten and first grade demonstrated relatively less cooperative participation (accepted the teacher's authority, complied with classroom rules). Children with whom teachers claimed to have relatively more conflictual relationships were thus at greater risk for poor academic achievement.

We did not expect teacher – child relationships to account for the entire association between aggression and achievement, given substantial evidence supporting Hinshaw's (1992) "third-variable" explanation. Many studies have shown that aggression and conduct disorders are strongly associated with hyperactivity and attention/distractibility problems (Connor et al., 2003; Fergusson & Horwood, 1995; Frick et al., 1991; Johnson, McGue, & Iacono, 2005; Spira & Fischel, 2005), which are also associated with poor academic achievement or underachievement (Frick et al., 1991; Hinshaw, 1992; Horn & Packard, 1985; Rabiner, Coie, & the Conduct Problems Prevention Research Group, 2000; Spira & Fischel, 2005; Velting & Whitehurst, 1997). Attention problems are not viewed as a mediator of the effects of aggressive behavior on achievement, but researchers who have attempted to separate their effects conclude that what appears to be an effect of conduct behaviors on achievement in elementary school can be explained substantially by the comorbidity of conduct disorders and hyperactivity/attention/distraction problems (Fergusson & Horwood, 1995; Frick et al., 1991; Hinshaw, 1992; Johnson et al., 2005). The independent effect of conduct on achievement appears to increase, however, in adolescence (Hinshaw, 1992). Most of these studies are based on clinical populations—children who have been diagnosed with attention-deficit hyperactive disorder, conduct disorders, and underachievement (a significant difference between actual achievement and achievement that would be predicted by IQ)—in contrast to the current study, which included mostly children who had no such diagnosis. This body of evidence is mentioned, however, because it reminds us that there are most likely a variety of valid explanations for associations between conduct and achievement and evidence for one explanation does not preclude the validity of another.

Children in this study were followed from kindergarten or first through fifth grades. Although we expected conflictual student – teacher relationships to mediate the effects of aggression on engagement and academic achievement for all these elementary grades, we took advantage of the longitudinal data to assess possible changes in the nature of the relationships over time. We also assessed gender differences, to determine whether boys were more prone to developing negative relationships that could undermine their engagement and learning, as is suggested by some previous studies discussed in the following, and to assess possible differences between boys and girls in the pattern of associations over time among the variables studied.

Gender Differences

Many previous studies have found girls to exhibit more prosocial and less antisocial behavior, and to have closer and less conflictual relationships with their teachers (e.g., Birch & Ladd, 1997; Bracken & Crain, 1994; Furrer & Skinner, 2003; Hamre & Pianta, 2001; Howes et al., 2000; Hughes, Cavell, & Willson, 2001; Kesner, 2000; Ryan, Stiller, & Lynch, 1994; Saft & Pianta, 2001; Silver et al., 2005). As noted by Hamre and Pianta (2005), because of the high percentage of female teachers, it is difficult to determine whether girls' advantage is related to the greater likelihood that their teacher is the same gender. The present study, like most, included too few male teachers to assess the effects of the teacher gender, but the design did allow us to assess gender differences in children from kindergarten or first grade through the fifth grade on these dimensions. Moderator effects of gender were also assessed in the longitudinal analyses, although we had no basis for making specific hypotheses.

The children in the study were ethnically diverse and came from very low-income families. This population was studied because although low-income children and children of color have, on average, the poorest achievement outcomes, most of the research on social skills and relationships in school involve middle-class samples. An understanding of the way social variables affect achievement in the children who are most at risk of school failure could help guide interventions designed to close the achievement gap. Previous studies have not focused on low-income children who are most at risk of poor performance, few have followed children for as long as 6 years through elementary school, and none have examined simultaneously direct effects of children's aggression on academic engagement and the effects of aggression on achievement mediated by relationships with teachers.

Method

Participants

Children. Participants of the original sample included 403 children (51.7% girls) who were participating in a longitudinal study on transition to school, although for most analyses, given missing data and attrition, there were approximately 300 children. Children were followed from kindergarten ($N = 283$) or first grade ($N = 120$) through the fifth grade. Some of the children had already completed kindergarten when the study began and thus could not begin their participation until the first grade. Kindergarteners and first graders were combined for analyses. The children who entered the study at either kindergarten or first grade were reassessed in second, third, and fifth grades. Analyses at each grade level included all the children in the longitudinal sample for whom relevant data were available at that grade level.

The families of all participating children had incomes below the federal poverty guidelines when they entered the study. At the start of the present study, most (76%) household incomes were below $15,000, 21% were below $6,000, most (62%) mothers were not married, and 49% were not working. With regard to education, 13% of the primary caretakers interviewed had less than the equivalent of a ninth-grade education and 28% went past the ninth grade but did not complete high school.

Of the participating children, 35% are African American, 34% are White, 28% are Latino, 2% are Asian, and 1% is Native American. Approximately 38% of the Latino children were not proficient in English at the beginning of the study. Children lived in three geographical areas of the United States: a rural and an urban area in the Northeast and a West Coast urban area.

Overall, attrition in the study was modest (less than 19% over a period of 6 years). Analyses comparing children who were lost to those who were still in the sample at fifth grade revealed no significant differences in family income, gender, or cognitive skills (measured at 60 months, before the present study began). A chi-square analysis of ethnic differences in attrition rates was, however, significant, $\chi^2(4) = 18.62$, $p \leq .001$. A higher proportion of Latino children were lost than either African American or Caucasian children.

Teachers. All teachers who had a study child enrolled in their classroom were invited to participate in the study. By the fifth grade, children were enrolled in 233 classrooms in 138 schools. The modal number of study children per classroom was 1. Teacher experience varied from 1 to 38 years of teaching, with a mean of 16 years. Almost all teachers (99%) had at least a bachelor's degree; 45% had a master's degree.

Procedures and Measures

Teacher questionnaires. In the spring of each year, questionnaires were mailed or delivered to teachers of participating children. Teachers responded to questions regarding an individual child's aggressive behavior, academic engagement, and teacher–child conflict. The average return rate for questionnaires was 75%.

Aggressive behavior. Teachers rated children's aggression using a subscale of the Child Behavior Scale, developed by Ladd and Profilet (1996). Four items (e.g., "fights with other children," "aggressive child," "taunts/teases other children," "bullies other children") were used, with a 3-point response format (*does not apply, applies sometimes, certainly applies*). Alpha coefficients were .91, .92, and .91 for kindergarten/first, third, and fifth, grades, respectively.

Teacher–child conflict. Teachers rated the degree to which they had a relationship that involved conflict for the individual child at each grade level using items from the conflict subscale of Pianta's Student–Teacher Relationship Scale (Pianta, 1994; Pianta & Steinberg, 1992; Pianta, Steinberg, & Rollins, 1995). Each item has a 5-point response scale (*definitely does not apply, not really, neutral/not sure, applies somewhat, definitely applies*). In consultation with the scale's author, 5 of the original 12 items from the conflict subscale were selected to diminish the burden on teachers (e.g., "this child and I always seem to be struggling with each other," "this child easily becomes angry with me"). The alphas for this scale were .89, .89, .92, and .88 for kindergarten/first, second, third, and fifth, grades, respectively.

Academic engagement. Teachers rated children's academic engagement—the degree to which an individual student is responsibly engaged with academic work—using items from the Teacher Rating Scale of School Adjustment (TRSSA) developed by Birch and Ladd (1997). Three of the items were taken from the self-directedness subscale of the TRSSA (e.g., "seeks challenges," "works independently"), and one of the items was from the cooperative participation subscale ("accepts responsibility for a given task") of the TRSSA. The items all have a 3-point response format (*does not apply, applies sometimes, and certainly applies*). Reliabilities for these four items were acceptable for each grade level (alphas = .87, .84, .83, and .85 for kindergarten/first, second, third, and fifth, grades, respectively).

Academic achievement. Children's math and literacy skills were assessed individually in the spring of kindergarten or first grade, and again in third and fifth grades, by a trained experimenter. The literacy and most of the math assessments were available in Spanish. The sections of the math assessment that were not available in Spanish were translated. Children who preferred the Spanish version were assessed by a native Spanish speaker. The number of children who were assessed in Spanish ranged from 38 (kindergarten/first) to 0 (fifth grade).

To ensure a range of scores and a broad array of math skills, kindergarteners and first graders were given a math test that included items from the Woodcock–Johnson Psychoeducational Battery–Revised (WJ–R) calculations subtest (Woodcock & Johnson, 1989), identification of written numbers and shapes from the Peabody Individual Achievement Test–Revised (Dunn & Dunn, 1981), word problems involving addition and subtraction developed by cognitively guided instruction (CGI) researchers (Carpenter, Ansell, Franke, & Fennema, 1993; Carpenter, Fennema, & Franke, 1996), questions about relative size (e.g., "which is more, 5 or 6?") from the Test of Early Mathematics Ability (Ginsburg & Baroody, 1990), and set of counting items (e.g., counting 50 blocks, counting from 90 to 120, and counting up or down from a specified number; Secada, 1991). Scores for each of these sets of items reflected the proportion of items answered correctly. The third- and fifth-grade math assessments included two subtests from the WJ–R: calculation and applied problem solving. Scores were calculated according to the manual instructions—the item number of the most difficult item answered correctly before missing five problems in a row minus the number of items missed since the basal item (based on the lowest of five items the student answered correctly). In addition to these two WJ–R subtests, children were given additional word problems taken from CGI research projects (Carpenter et al., 1993; Carpenter et al., 1996). Third graders received one additional set of CGI problems that involved counting using individual and sticks of 10 unifix cubes. The total math achievement score was calculated by standardizing the score for each subtest within grade and averaging them. Alpha reliability coefficients ranged from .55 (third grade) to .80, with an average of .71. We suspect that the relatively low alpha in math for third graders is a consequence of items (e.g., using unifix cubes) that were unfamiliar to some children.

For literacy achievement, two subtests of the WJ–R were used for the third and fifth graders: the letter–word reading and the passage comprehension. The first test asks children to decode single words. In the second, children read a passage and fill in a missing word. The WJ–R letter–word identification subscale was adapted for the kindergarten and first graders to ensure enough items at their level to show some variability in scores. (Pilot testing revealed a floor effect.) Thus, instead of asking children to identify only the 9 letters in the WJ–R, they were shown all 26 letters of the alphabet. Kindergarten and first-grade children were also given a larger number of simple (2 to 4 letters) words to identify than are provided in the WJ–R. The WJ–R was supplemented at all grades with the Saunders (1999) performance-based test, which assesses writing, comprehension, and reading and verbal fluency. Children were dictated letters (kindergarten only), words, and sentences to write. Scores reflected the proportion of letters or words written correctly. Comprehension was assessed by reading a passage to the child (kindergarten) or having a child read a passage (first, third, and fifth grades), and asking a series of questions about the story. Answers were rated on a 1–3 scale for their correctness, and these scores were averaged to create a comprehension score. Overall verbal fluency was rated on a 0 (practically nonverbal) to 2 (verbal, fluent, talkative; gave detailed answers) scale. Reading fluency was assessed on a 0 (nonreader) to 6 (reads words fluently) scale. All subscale scores were converted to standardized scores and then combined within each grade to create an overall mean literacy achievement score. Alphas for the literacy scores at each grade ranged from .74 to .82, with an average of .79.

Assessments of children's math and literacy skills were highly correlated at every grade (.58, .55, and .55 for kindergarten/first, third, and fifth grades, respectively, all $ps \leq .001$). Because we had no reason to expect aggression to affect or be affected differently by math and literacy skills, standardized scores for the two subject areas were combined to form one measure of academic achievement. The alphas for this combined achievement measure were .73, .72, and .71 for kindergarten/first, third, and fifth grades, respectively.

Results

The percentage of children missing data on the variables included in this study ranged from 8% to

24% of the sample. Children with missing data did not differ significantly from children without missing data on ethnicity or gender, with one exception. Latino children were more likely to have missing teacher-reported data in Grades 1, 3, and 5 than White and African American children. Because hierarchical linear modeling (HLM) allows for ignorably missing data through its use of weighting and smoothing techniques, no imputations were made and all data were used in the growth curve analysis. For the path analysis, only the participants who had complete data at each grade level were included.

Table 1 presents the means and standard deviations, by gender, for all variables used in analyses. The distributions for aggression were positively skewed, with Pearson skewness coefficients ranging from 1.17 to 1.34. The distribution of teacher – child conflict ratings were also positively skewed, with Pearson skewness coefficients ranging from .98 to 1.41. Log transformation of these variables did not affect the findings.

Interconstruct and Intragrade Correlations

As a first step in testing the hypotheses, correlations among all the variables of interest were computed (see Table 2). As expected, and consistent with previous research (e.g., Birch & Ladd, 1998), children's aggression ratings were generally positively associated with teacher – child conflict and negatively associated with academic engagement and achievement. Teacher – child conflict, in turn, was generally negatively associated with academic engagement and achievement, and academic engagement was significantly correlated with academic achievement. The one notable anomaly was that academic achievement was not associated with either aggression or teacher – child conflict at third grade. Intergrade correlations shown in Table 3 indicate a fair amount of stability in all the variables studied.

Explanations for Associations Between Aggression and Achievement

Growth curve analyses. Growth curve analysis using HLM 6.21 (Raudenbush & Bryk, 2002) was used to test three sets of hypotheses bearing on different explanations for the association between aggression and achievement: (a) that changes in achievement predict changes in aggression, (b) that changes in aggression predict changes in achievement, and (c) that the effect of aggression on achievement is mediated by teacher – student relationships. To assess the mediation hypothesis we first tested each of the proposed links in the model: (a) changes in aggression predict changes in teacher – child conflict, (b) changes in teacher – child conflict predict changes in engagement, which (c) in turn predict changes in academic achievement. We then tested the mediation hypothesis more directly by assessing the hypothesis that (d) increases in aggression no longer predict changes in achievement with mean teacher – child conflict (and in a subsequent analysis, mean engagement) held constant.

HLM was used because it estimates individual and group growth curves simultaneously to describe patterns of change over time. The individual growth curves can be predicted from time-varying variables as well as testing mediation and moderation effects of between-subjects and time-invariant factors, such as gender and mean teacher – child conflict, and HLM can accommodate missing values.

Table 1
Means and Standard Deviations for Aggression, Teacher – Child Conflict, Engagement, and Achievement, by Gender

	Aggression	Conflict	Engagement	Achievement
Kindergarten/first grade				
Boys (n = 151)	1.55 (.55)	1.89 (0.98)	2.28 (.51)	−0.02 (0.66)
Girls (n = 150)	1.37 (.42)	1.70 (0.91)	2.40 (.56)	0.03 (0.73)
Second grade				
Boys (n = 162)	1.57 (.64)	2.04 (1.00)	2.07 (.53)	—
Girls (n = 168)	1.36 (.42)	1.74 (0.78)	2.26 (.40)	—
Third grade				
Boys (n = 162)	1.54 (.53)	2.00 (1.00)	2.11 (.46)	−0.03 (0.65)
Girls (n = 166)	1.39 (.49)	1.71 (0.91)	2.20 (.52)	0.01 (0.60)
Fifth grade				
Boys (n = 137)	1.58 (.60)	2.16 (0.96)	2.04 (.51)	−0.09 (1.34)
Girls (n = 143)	1.27 (.47)	1.67 (0.89)	2.24 (.52)	0.14 (1.29)

Table 2
Correlations Among Constructs, Within Grades

	Conflict	Engagement	Achievement
Kindergarten/first grade			
Aggression	.48***	−.29**	−.14**
Teacher–child conflict		−.52***	−.18**
Engagement			.44***
Second grade			
Aggression	.70***	−.35***	—
Teacher–child conflict		−.46***	—
Engagement			—
Third grade			
Aggression	.65***	−.26***	−.07
Teacher–child conflict		−.40***	−.02
Engagement			.30***
Fifth grade			
Aggression	.20**	−.29**	−.24**
Teacher–child conflict		−.45***	−.20**
Engagement			.30***

Note. The sample sizes were 301 for kindergarten/first grade, 330 for second grade, 228 for third grade, 280 for fifth grade; no achievement test was given in second grade.
$^{**}p \leq .01$. $^{***}p \leq .001$.

For each model tested, the Level 1 model looked as such,

$$y_{it} = \pi_{0i} + \pi_{1i} \times (\text{Grade}_{it}) + \pi_{2i} \times (\text{Grade} \times \text{Predictor}_{it}) + e_{it}.$$

The term y_{it} is the outcome variable for child i at time t and is a linear function of the child's grade in school on that occasion plus random error (e_{it}). The equation stipulates that the trajectory is linear across time and has the growth parameters: π_{0i}, π_{1i}, and π_{2i}. A linear model was used because we had just three time points.

The intercept π_{0i} should be interpreted as each child's mean score of the outcome variable at the initial time, which is in either kindergarten or first grade. The second parameter, π_{1i}, represents the slope of the outcome variable or the rate the individual child changes across the grades (coded as *kindergarten or first grade* = 1, *third grade* = 3, and *fifth grade* = 5). Second grade was not included in this analysis because an independent achievement assessment was not given in second grade. The third parameter, π_{2i}, represents the rate of change on the outcome variable over time, taking into account change in the predictor. This parameter tests whether changes in the predictor across grades is associated with changes in the outcome variable. The random error for child i at time t is represented in the equation as e_{it}. This may be interpreted as the measurement errors associated with the differences between the true and observed growth trajectories.

The first model (Model 1) tested whether changes in achievement predicted changes in aggression. As seen in Table 4, the estimation of fixed effects indicates that change over time in children's achievement was a marginally significant predictor of change over time in children's aggression ($\beta = -.01$, $p \leq .10$); on average, an increase in achievement was associated with a decrease in aggression.

The second model (Model 2) tested whether changes in aggression across the grades predicted changes in achievement. The results (Table 4) indicate that on average an increase in aggression was significantly associated with a decrease in achievement ($\beta = -.09$, $p \leq .05$).

The third set of models (Models 3a, 3b, 3c) tested whether (a) changes in aggression predicted changes in teacher–child conflict, (b) changes in teacher–child conflict predicted changes in engagement, and (c) changes in engagement predicted changes in children's achievement (see Table 4). The results indicate that changes in aggression significantly predicted changes in teacher–child conflict; an increase in aggression was associated with an increase in conflict ($\beta = .31$, $p \leq .001$). In addition, an increase in conflict significantly predicted a decrease in engagement ($\beta = -.07$, $p \leq .001$). Last, an increase in engagement was significantly associated with an increase in achievement ($\beta = .19$ $p \leq .001$).

Table 3

Intergrade Correlations

	Second grade	Third grade	Fifth grade
Aggression			
Kindergarten/first grade	.54***	.32**	.48***
Second grade		.27***	.56***
Third grade			.35***
Teacher–child conflict			
Kindergarten/first grade	.47***	.39***	.28***
Second grade		.43***	.33***
Third grade			.37***
Engagement			
Kindergarten/first grade	.33***	.29***	.35***
Second grade		.45***	.16*
Third grade			.32***
Achievement			
Kindergarten/first grade	—	.65***	.58***
Third grade			.71***

Note. The sample sizes are 301 for kindergarten/first grade, 330 for second grade, 228 for third grade, 280 for fifth grade; no achievement test was given in second grade.
$^*p \leq .05$. $^{**}p \leq .01$. $^{***}p \leq .001$.

To assess the goodness of fit of the models tested, a comparison of the deviance and the chi-square statistics among models was conducted. The deviance statistic of a full maximum likelihood model reflects the fit of the entire model and signifies how much worse the current model is when compared to the best possible model. Thus, the smaller the absolute value is, the closer the fit to the best possible model (Singer & Willett, 2003). For Models 1–3, the deviance statistics were as follows: Model 1 = 2,502.23, Model 2 = 1,577.92, Model 3a = 2,683.24, Model 3b = 1,822.39, and Model 3c = 4,171.74. The chi-square statistics also indicate the relative fit of a model to the data. In the present study, none of the models' chi-square statistics were significant (all $ps \leq .50$). Thus, we are able to reject the null hypotheses for all models, which is desired when testing the fit of a model.

Gender differences. To assess the role of gender (coded as 1 = boys, 0 = girls) in the growth curve analyses, gender was included in the earlier models at the second level, such that:

$$\pi_{0i} = \beta_{00} + \beta_{01} \times (\text{gender}) + r_0$$

$$\pi_{1i} = \beta_{10} + \beta_{11} \times (\text{gender})$$

$$\pi_{2i} = \beta_{20} + \beta_{21} \times (\text{gender})$$

Results (see Table 4) indicate that gender significantly predicted initial differences in aggression, teacher–child conflict, and student engagement; on average, boys had significantly higher ratings of aggression ($\beta = .13, p \leq .05$) and teacher–child conflict ($\beta = .22, p \leq .05$), and, initially, girls had significantly higher engagement ($\beta = -.16, p \leq .01$). The test of gender differences in initial achievement was not significant. Gender did not show any significant effects on changes in these four variables across grades and gender did not moderate the effect of the predictor on the outcome in any analysis (see Table 4).

Mediational models. Finally, we tested directly whether mean teacher–child conflict or student engagement mediated the relationship between aggression and achievement. Mean teacher–child conflict and student engagement were added into the Level 2 model separately such that:

$$\begin{aligned}\pi_{0i} = {} & \beta_{00} + \beta_{01} \\ & \times (\text{mean teacher–child conflict or engagement}) \\ & + r_0 \\ \pi_{1i} = {} & \beta_{10} + \beta_{11} \\ & \times (\text{mean teacher–child conflict or engagement}) \\ \pi_{2i} = {} & \beta_{20} + \beta_{21} \\ & \times (\text{mean teacher–child conflict or engagement})\end{aligned}$$

As shown in Table 5, the results indicate that neither mean teacher–child conflict ($\beta = .05, p = .31$) nor mean student engagement ($\beta = .04, p = .50$) significantly mediated the effect of changes in aggression on changes in achievement.

Table 4
Fixed Effects of Linear Growth Curve Analyses

Fixed effect (with robust *SE*s)	Coefficient	*SE*	*t* ratio	*p* value
Model 1, achievement predicting aggression				
Intercept	1.42	.04	39.97	.00
Gender	0.13	.06	2.25	.03
Slope	−0.02	.01	−2.12	.03
Gender	0.02	.02	1.54	.12
Change in achievement	−0.01	.00	−1.82	.07
Gender	0.00	.00	0.10	.92
Model 2, aggression predicting achievement				
Intercept	1.47	.07	19.70	.00
Gender	−0.02	.11	−0.19	.85
Slope	−0.11	.05	−2.05	.04
Gender	−0.01	.08	−0.15	.88
Change in aggression	−0.09	.04	−2.21	.03
Gender	0.00	.05	0.08	.94
Model 3a, aggression predicting teacher–child conflict				
Intercept	1.65	.07	23.21	.00
Gender	0.22	.11	2.00	.05
Slope	−0.39	.04	−9.80	.00
Gender	0.03	.05	0.49	.62
Change in aggression	0.31	.03	11.30	.00
Gender	−0.05	.04	−1.28	.20
Model 3b, teacher–child conflict predicting engagement				
Intercept	2.37	.04	59.37	.00
Gender	−0.16	.06	−2.65	.01
Slope	0.07	.02	4.58	.00
Gender	0.01	.02	0.35	.73
Change in teacher–child conflict	−0.07	.01	−8.99	.00
Gender	0.01	.01	0.90	.37
Model 3c, engagement predicting achievement				
Intercept	1.42	.07	20.72	.00
Gender	−0.01	.10	−0.12	.91
Slope	−0.64	.08	−8.12	.00
Gender	0.12	.11	1.12	.27
Change in engagement	0.19	.03	5.86	.00
Gender	−0.06	.05	−1.14	.26
Model 4, teacher–child conflict predicting aggression				
Intercept	1.40	.03	43.94	.00
Gender	0.17	.06	3.16	.00
Slope	−0.16	.01	−11.45	.00
Gender	−0.04	.02	−1.57	.12
Change in teacher–child conflict	0.08	.01	10.23	.00
Gender	0.01	.01	0.83	.41

Within-grade path analyses. The growth curve analyses have several limitations. They do not allow a test of the full hypothesized causal model (aggression → teacher–child conflict → low engagement → low achievement). Second, because there are only three time points, it cannot be determined whether the significant effect is primarily based on predictions from kindergarten/first to third grade or third grade to fifth grade. A related limitation is that they take good advantage of the longitudinal data, but they do not tell us whether the hypothesis is supported within each grade.

Path analyses were therefore conducted to test the mediation hypothesis at individual grade levels. In the model was a path that included the mediators (teacher–child conflict and engagement) as well as a direct path from aggression to academic achievement. To assess the overall fit of each model at

Table 5
Mediational Models for Aggression Predicting Achievement

Fixed effects (with robust *SE*s)	Coefficient	*SE*	*t* ratio	*p* value
Intercept	1.46	.16	9.26	.00
Mean teacher – child conflict	−0.00	.08	−0.02	.99
Slope	0.06	.13	0.49	.62
Mean teacher – child conflict	−0.11	.07	−1.61	.11
Change in aggression	−0.17	.11	−1.60	.11
Mean teacher – child conflict	0.05	.05	1.02	.31
Intercept	1.26	.30	4.18	.00
Mean engagement	0.08	.14	0.59	.56
Slope	−0.70	.20	−3.41	.00
Mean engagement	0.22	.09	2.44	.02
Change in aggression	−0.10	.12	−0.82	.41
Mean engagement	0.04	.06	0.68	.50

each grade level, LISREL 8 (Jöreskog & Sörbom, 1993) was used. In addition, all path coefficients in Figures 1a – 1c are standardized beta coefficients.

Figure 1a shows the results of the path analysis for kindergarten/first grade. The results from the various tests of model fit allow us to conclude that the specified kindergarten/first-grade model fit the data well. The chi-square test ($\chi^2 = 2.68$, $p = 0.26$) was not statistically significant; that is, we failed to reject the null hypothesis, which is desired in testing a model. The root mean square error of approximation (RMSEA) was 0.03. Byren (1998) notes that a model with RMSEA values less than 0.05 indicates a good fit, whereas values ranging from 0.08 to 0.10 indicate a mediocre fit. The goodness of fit index (GFI), in which 0.9 indicates a relatively good fit, was 0.98. The results from the individual paths show that the paths from aggression to teacher – child conflict, conflict to engagement, and engagement to achievement are significant but the direct path from aggression to achievement is not significant.

Figure 1b shows the results of the path analysis for third grade. The results from the various tests of model fit allow us to conclude that the specified third-grade model fit the data well. The chi-square test ($\chi^2 = 3.54$, $p = 0.17$) was not statistically significant and the RMSEA was 0.05. The GFI was 0.97. Similar to the results in kindergarten/first grade, the paths from aggression to conflict, conflict to engagement, and engagement to achievement were significant, whereas the direct path from aggression to achievement was not.

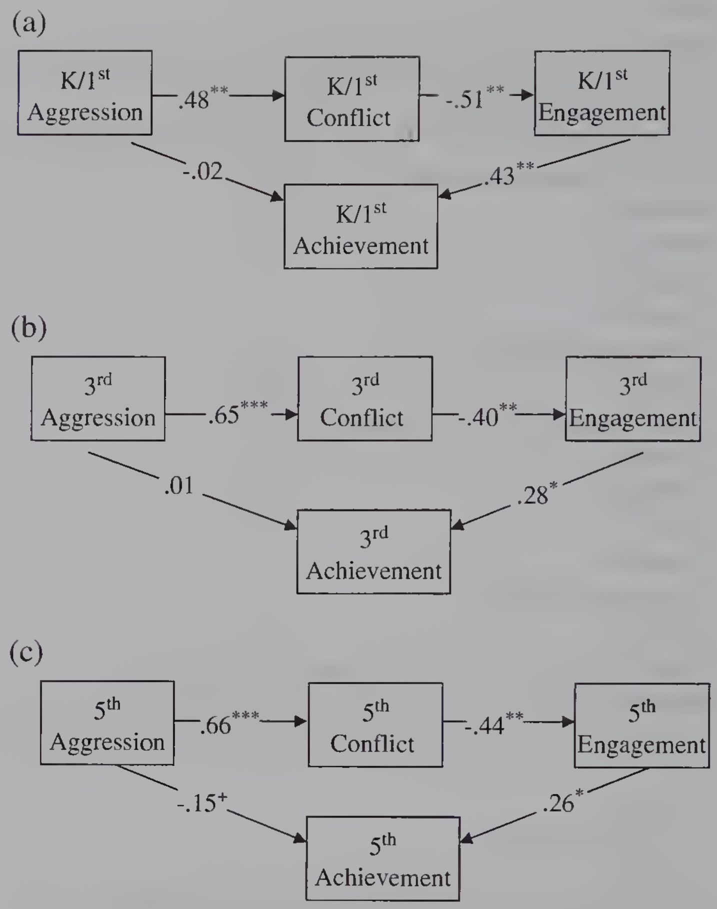


Figure 1. Path analysis results, predicting achievement within grade: (a) kindergarten/first grade, (b) third grade, and (c) fifth grade. $^{*}p \leq .05$. $^{**}p \leq .01$. $^{***}p \leq .001$.

Figure 1c shows the results of the path analysis for fifth grade. The results from the various tests of model fit allow us to conclude that the specified fifth-grade model fit the data fairly well. The chi-square test ($\chi^2 = 1.41$, $p = 0.52$) was not statistically significant. The RMSEA of 0.09 indicated a mediocre fit and the GFI was .95. All the paths representing the mediation hypothesis had significant standardized beta coefficients. Unlike the findings for the younger grades, in which the direct path from aggression to achievement was not significant, for fifth grade it was marginally significant.

In summary, the path analyses are consistent with the hypothesis that the effect of aggression on academic achievement is at least partially mediated by its effects on conflict with teachers and the resultant lowered engagement in academic work. The direct path between aggression and achievement was not significant, although by fifth grade, it was almost significant.

Predicting Aggression From Teacher – Child Conflict

As discussed earlier, there is evidence suggesting that teacher – child relationships are both affected by and affect children's social behavior. In the aforementioned analyses, teacher – child conflict was predicted by children's aggression. In the next analysis we assess evidence for a relationship in the opposite direction. Results of a growth curve analysis (see Table 4, Model 4) indicate that changes in teacher – child conflict significantly predicted changes in aggression; an increase in conflict across the grades was associated with an increase in aggression (β = .08, $p \leq .001$). The deviance statistic to test the goodness of fit for Model 4 was 1,250.31 and the chi-square statistic measuring the homogeneity of the Level 1 variance was nonsignificant, suggesting a good fit.

To assess whether the effect of teacher–child conflict on the change in aggression was different for boys than girls, gender (coded as 1 = *boys*, 0 = *girls*) was included in Model 4 at the second level. The results (see Model 4, Table 4) of the main effect of gender on the change in aggression indicate that gender significantly predicted initial differences in aggression, with boys, on average, having significantly higher ratings in aggression (β = .17, $p \leq .01$). However, the test of gender differences in change in aggression was not significant (β = −.04, $p = .12$).

Discussion

This study examined several alternative, although not mutually exclusive, explanations for the association often found between aggression and achievement (see Duncan et al., 2007, for an exception in that it found few significant associations between social behavior and school performance). The study also assessed gender differences on these dimensions and in the pattern of relationships among them. These topics are discussed in the following.

Explanation of Aggression–Achievement Association

The significant cross-grade correlations are consistent with previous studies suggesting that children's social behavior remains relatively stable over the first few grades of school (e.g., Birch & Ladd, 1998; Ladd & Burgess, 2001; Ladd & Profilet, 1996; Silver et al., 2005). Nevertheless, changes that were seen in children's aggression predicted changes in achievement, suggesting that aggression may undermine achievement. There was also some, albeit weaker, evidence for the view that poor academic achievement promotes aggression.

Why might academic skills affect children's behavior? We have previously proposed that children who find schoolwork difficult may become frustrated or unhappy in school and express their frustration and unhappiness by acting aggressively (Miles & Stipek, 2006). Our proposal is based on an early theory that frustration (defined as the state that emerges when circumstances interfere with a goal response) leads to aggression (Dollard, Doob, Miller, Mowrer, & Sears, 1939). Berkowitz (1968) suggested that any unpleasant emotion and feeling, such as anxiety, anger, and annoyance, can trigger either "fight or flight." There is some empirical evidence for the theory (see Berkowitz, 1978; Buss, 1963). Rutter and Guller (1983) proposed a somewhat different but related explanation for the effect of poor achievement on aggression. They suggested that academic failure leads to low self-esteem and general antagonism toward school, which in turn prompts negative behaviors.

By fifth grade, there was some weak evidence for a direct effect of aggression on achievement. But most of the findings were consistent with our hypothesis that the effect of aggression on achievement is partially mediated by its effects on teacher–child relationships. The growth curve analyses supported each link in the model, showing that changes in aggression significantly predicted changes in the amount of teacher–child conflict, which in turn predicted changes in students' engagement, which predicted changes in students' achievement. Although gender was associated with mean initial scores on all but the achievement measure, gender did not moderate the associations found among the variables. The mediation hypothesis was also supported by path analyses within each grade evaluating the entire model, although not in the growth curve analysis. The failure of the test of the mediators in the growth curve analyses may have occurred because only the average of the mediator over the three grades could be tested, not change over time in the mediators, and teacher–child conflict and engagement had to be tested in separate analyses. Although the growth curve analysis took advantage of the longitudinal data, it was not the most effective way to assess the full mediation model.

There are many reasons why developing a conflictual relationship with the teacher might undermine academic engagement and learning. For example, children who have relatively high amounts of conflict with the teacher may develop generally negative feelings about the teacher and school, and as a result lack motivation to engage in the academic work the teacher assigns. Teachers may also report a conflictual relationship with the children they discipline relatively often; the frequent experience of being disciplined may contribute to the children's negative feelings as well as take time away from academic work.

The results of this study highlight the importance of helping children develop self-regulation, anger management, and social problem-solving skills. Even if academic performance is the primary concern, there may be value in investing time and effort in children's social experiences and skills. Positive social skills and relationships might serve as important protective mechanisms against academic failure, which might be particularly important for low-income children, who are often found to be at risk for poor academic performance.

Relationships with teachers and engagement, as measured in this study, did not completely explain the effects of aggression on achievement. In the longitudinal analyses changes in aggression continued to predict changes in achievement, with mean teacher–child conflict and with mean engagement held constant. And for fifth graders, the direct path from aggression to achievement was marginally significant. Also, despite the consistently strong associations found between aggression and achievement, the results of this study do not rule out the possibility that aggression is not really the culprit but rather other problems that co-occur with aggression undermine academic achievement. Previous research, mentioned earlier, suggests that hyperactivity and accompanying attention problems are really at fault, not aggression. Most previous studies on hyperactivity and attention problems included children who were diagnosed with significant problems on these dimensions. It would be useful in future studies to assess the independent effects of attention/hyperactivity and aggression on children's learning, both directly and indirectly, for children who cover a broader range on these dimensions.

Effect of Teacher–Child Conflict on Aggression

The study also examined whether teacher–child conflict predicted aggression, consistent with previous research, summarized earlier (e.g., Hamre & Pianta, 2001; Silver et al., 2005), suggesting that teacher–child relationships may promote or undermine positive behavior. As predicted, not only did changes in aggression predict changes in teacher–child conflict but the reverse was also true. Changes in teacher–child conflict also predicted changes in children's aggression. The findings are consistent, therefore, with a reciprocal relationship, with aggressive behavior generating conflict with teachers and conflict with teachers exacerbating children's aggressive behavior.

The findings that teacher–child conflict predicted both increased aggression and lower engagement suggest the importance of helping teachers avoid developing conflictual relationships with aggressive children. We recognize that this may be difficult, given that teacher behavior is in part a reaction to child behavior. But a preponderance of negative interactions with the teacher is not likely to support children's desire to engage in teacher-sanctioned activities. Efforts to ensure at least some positive interactions, for example, by engaging the more aggressive children in conversation about their interests, taking care to notice good behavior, and conveying high expectations, may compensate for the negative disciplinary interactions.

Gender Differences

As hypothesized, already by kindergarten or first grade, teachers rated boys as more aggressive and less engaged, and their relationships with girls as less conflictual. These findings are consistent with previous research (e.g., Birch & Ladd, 1997; Bracken & Crain, 1994; Furrer & Skinner, 2003; Hamre & Pianta, 2001; Howes et al., 2000; Hughes et al., 2001; Kesner, 2000; Ryan et al., 1994; Saft & Pianta, 2001; Silver et al., 2005) and have important practical implications. In light of evidence that social competence, teacher–child relationships, and academic engagement affect children's academic learning, the present study provides further evidence that boys are at greater risk of underachievement than girls. Given the possibility of a key role played by relationships with teachers in children's engagement and achievement, teachers might need to make extra efforts to develop positive relationships with boys, especially boys who are aggressive. Such efforts might serve as a protective factor and improve boys' academic success.

Limitations and Conclusions

This study relied substantially on teacher ratings. It would be useful in future studies to obtain independent assessments of children's aggression, relationships with teachers, and engagement in academic activities. Children's own perceptions related to all of these constructs would also be useful. Furthermore, data were gathered at one time point at each grade level. Future work could examine children's social competencies, relationships with teachers, and achievement at both the beginning and the end of each school year.

Despite some limitations, the study reminds us of the complexity of relationships among variables that affect student achievement. Many factors affect children's learning. This study, combined with previous research, suggests that even one factor, aggressive behavior, probably affects achievement in more than one way and that the relationship between aggression and achievement is very likely reciprocal. Developing analytic methods to assess these complex relationships is a challenge but an important goal of developmental research.

References

Arnold, D. (1997). Co-occurrence of externalizing behavior problems and emergent academic difficulties in young high-risk boys: A preliminary evaluation of patterns and mechanisms. *Journal of Applied Developmental Psychology, 18*, 317–330.

Berkowitz, L. (1968). The study of urban violence: Some implications of laboratory studies of frustration and aggression. *American Behavioral Scientist, 11*(4), 14–17.

Berkowitz, L. (1978). Whatever happened to the frustration-aggression hypothesis? *American Behavioral Scientist, 21*(5), 691–708.

Birch, S., & Ladd, G. (1997). The teacher-child relationship and children's early school adjustment. *Journal of School Psychology, 35*, 61–79.

Birch, S., & Ladd, G. (1998). Children's interpersonal behaviors and the teacher-child relationship. *Developmental Psychology, 34*, 934–946.

Bracken, B. A., & Crain, R. M. (1994). Children's and adolescents' interpersonal relations: Do age, race, and gender define normalcy? *Journal of Psychoeducational Assessment, 12*, 14–32.

Buss, A. (1963). Physical aggression in relation to different frustrations. *Journal of Abnormal & Social Psychology, 67*,1–7.

Byren, B. M. (1998). *Structural equation modeling with LISERL, PRELIS and SIMPLIS: Basic concepts, applications, and programming*. Mahwah, NJ: Erlbaum.

Carpenter, T., Anseil, E., Franke, M., & Fennema, E. (1993). Models of problem solving: A study of kindergarten children's problem solving processes. *Journal for Research in Mathematics Education, 24*, 428–441.

Carpenter, T., Fennema, E., & Franke, M. (1996). Cognitively guided instruction: A knowledge base for reform in primary mathematics instruction. *Elementary School Journal, 97*, 3–20.

Coie, J. D., & Dodge, K. A. (1988). Multiple sources of data on social behavior and social status in the school: A cross-age comparison. *Child Development, 59*, 815–829.

Connell, J. (1990). Context, self and action: A motivational analysis of self-system processes across the life span. In D. Cicchetti & M. Beeghly (Eds.), *The self in transition: Infancy to childhood* (pp. 61–97). Chicago: University of Chicago Press.

Connell, J., & Wellborn, J. (1991). Competence, autonomy, and relatedness: A motivational analysis of self-system processes. In M. Gunnar & L. A. Sroufe (Eds.), *Self-processes and development: The Minnesota symposia on child development* (pp. 43–77). Hillsdale, NJ: Erlbaum.

Connor, D., Edwards, G., Fletcher, K., Baird, J., Barkley, R., & Steingard, R. (2003). Correlates of comorbid psychopathology in children with ADHD. *Journal of the American Academy of Child & Adolescent Psychiatry, 42*, 193–200.

Crick, N. R., Grotpeter, J., & Bigbee, M. (2002). Relationally and physically aggressive children's intent attributions and feelings of distress for relational and instrumental peer provocations. *Child Development, 73*, 1134–1142.

Crick, N. R., Werner, N. E., Casas, J. F., O'Brien, K. M., Nelson, D. A., Grotpeter, J. K., et al. (1999). Childhood aggression and gender: A new look at an old problem. In D. Bernstein (Ed.), *Nebraska symposium on motivation* (Vol. 45). Lincoln: University of Nebraska Press.

Crosnoe, R., Johnson, M. K., & Elder, G. H. (2004). Intergenerational bonding in school: The behavioral and contextual correlates of student–teacher relationships. *Sociology of Education, 77*, 60–81.

Deci, E., & Ryan, R. (1985). The general causality orientations scale: Self-determination in personality. *Journal of Research in Personality, 19*, 109–134.

Dollard, J., Doob, L. W., Miller, N. E., Mowrer, O. H., & Sears, R. R. (1939). *Frustration and aggression*. Oxford, England: Yale University Press.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., et al. (2007). *Developmental Psychology, 43*, 1428–1446.

Dunn, L. M., & Dunn, L. M. (1981). *Peabody Individual Achievement Test–Revised*. Circle Pines, MN: American Guidance Service.

Fergusson, D., & Horwood, L. (1995). Early disruptive behavior, IQ, and later school achievement and delinquent behavior. *Journal of Abnormal Child Psychology, 23*, 183–199.

Frick, P., Kamphaus, R., Lahey, B., Loeber, R., Christ, M., Hart, E., & Tannenbaum, L. (1991). Academic underachievement and the disruptive behavior disorders. *Journal of Consulting and Clinical Psychology, 59*, 289–294.

Furrer, C., & Skinner, E. (2003). Sense of relatedness as a factor in children's academic engagement and performance. *Journal of Educational Psychology, 95*, 148–162.

Ginsburg, H., & Baroody, A. (1990). *Test of early mathematics ability* (2nd ed.). Austin, TX: Pro-Ed.

Gregory, A., & Weinstein, R. S. (2004). Connection and regulation at home and in school: Predicting growth in achievement for adolescents. *Journal of Adolescent Research, 19*, 405–427.

Hamre, B. K., & Pianta, R. C. (2001). Early teacher-child relationships and the trajectory of children's school outcomes through eighth grade. *Child Development, 72*, 625–638.

Hamre, B. K., & Pianta, R. (2005). Can instructional and emotional support in the first-grade classroom make a difference for children at risk of school failure? *Child Development, 76*, 949–967.

Hinshaw, S. (1992). Externalizing behavior problems and academic underachievement in childhood and adolescence: Causal relationships and underlying mechanisms. *Psychological Bulletin, 111*, 127–155.

Horn, W., & Packard, T. (1985). Early identification of learning problems: A meta-analysis. *Journal of Educational Psychology, 77*, 597–607.

Howes, C., Phillipson, L., & Peisner-Feinberg, E. (2000). The consistency of perceived teacher-child relationships between preschool and kindergarten. *Journal of School Psychology, 38*, 113–132.

Hughes, J., Cavell, T., & Willson, V. (2001). Further support for the developmental significance of the quality of the teacher-student relationship. *Journal of School Psychology, 39*, 289 – 301.

Johnson, W., McGue, M., & Iacono, W. (2005). Disruptive behavior and school grades: Genetic and environmental relations in 11-year-olds. *Journal of Educational Psychology, 97*, 391 – 405.

Jöreskog, K., & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with SIMPLIS command language*. Hillsdale, NJ: Erlbaum.

Jorm, A. F., Share, D. L., Matthews, R., & Maclean, R. (1986). Behaviour problems in specific reading retarded and general reading backward children: A longitudinal study. *Journal of Child Psychology & Psychiatry & Allied Disciplines*, *27*, 33 – 43.

Kesner, J. (2000). Teacher characteristics and the quality of child-teacher relationships. *Journal of School Psychology, 28*, 133 – 149.

Ladd, G. W., Birch, S., & Buhs, E. (1999). Children's social lives in kindergarten: Related spheres of influence. *Child Development*, *70*, 1373 – 1400.

Ladd, G. W., & Burgess, K. (2001). Do relational risks and protective factors moderate the linkages between childhood aggression and early psychological and school adjustment? *Child Development*, *72*, 1579 – 1601.

Ladd, G. W., & Profilet, S. (1996). The child behavior scale: A teacher-report measure of young children's aggressive, withdrawn, and prosocial behaviors. *Developmental Psychology, 32*, 1008 – 1024.

Little, T. D., Brauner, J., Jones, S. M., Nock, M. K., & Hawley, P. H. (2003). Rethinking aggression: A typological examination of the functions of aggression. *Merrill-Palmer Quarterly, 49*, 343 – 367.

Lynch, M., & Cicchetti, D. (1997). Children's relationships with adults and peers: An examination of elementary and junior high school students. *Journal of School Psychology, 35*, 81 – 99.

Mantzicopoulos, P. (2005). Conflictual relationships between kindergarten children and their teachers: Associations with child and classroom context variables. *Journal of School Psychology, 43*, 425 – 442.

McGee, R., Williams, S., Share, D., Anderson, J., & Silva, P. (1986). The relationship between specific reading retardation, general reading backwardness and behavioural problems in a large sample of Dunedin boys: A longitudinal study from five to eleven years. *Journal of Child Psychology & Psychiatry, 27*, 597 – 610.

Miles, S. B., & Stipek, D. (2006). Contemporaneous and longitudinal associations between social behavior and literacy achievement in low-income elementary school children. *Child Development*, *77*, 103 – 117.

National Research Council Committee on Increasing High School Students' Engagement and Motivation to Learn. (2004). *Engaging schools: Fostering high school motivation to learn*. Washington, DC: National Academies Press.

Pianta, R. C. (1994). Patterns of relationships between children and kindergarten teachers. *Journal of School Psychology, 32*, 15 – 31.

Pianta, R. C., & Steinberg, M. (1992). Teacher-child relationships and the process of adjusting to school. *New Directions for Child Development*, *57*, 61 – 80.

Pianta, R. C., Steinberg, M., & Rollins, K. (1995). The first two years of school: Teacher-child relationships and deflections in children's classroom adjustment. *Development and Psychopathology, 7*, 295 – 312.

Pianta, R. C., & Stuhlman, M. W. (2004). Teacher-child relationships and children's success in the first years of school. *School Psychology Review, 33*, 444 – 458.

Rabiner, D., Coie, J. D., & the Conduct Problems Prevention Research Group. (2000). Early attention problems and children's reading achievement: A longitudinal investigation. *Journal of the American Academy of Child and Adolescent Psychiatry, 39*, 859 – 867.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Rutter, M., & Guller, H. (1983). *Juvenile delinquency: Trends and perspectives*. Suffolk, England: Penguin Books.

Ryan, R. M., & Deci, E. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well being. *American Psychologist*, *55*, 68 – 78.

Ryan, R. M., Stiller, J. D., & Lynch, J. H. (1994). Representations of relationships to teachers, parents, and friends as predictors of academic motivation and self-esteem. *Journal of Early Adolescence,14*(2), 226 – 249.

Saft, E., & Pianta, R. (2001). Teachers' perceptions of their relationships with students: Effects of child age, gender, and ethnicity of teachers and children. *School Psychology Quarterly, 16*, 125 – 141.

Saunders, W. (1999). Improving literacy achievement for English learners in transitional bilingual programs. *Educational Research and Evaluation*, *5*, 345 – 381.

Secada, W. G. (1991). Degree of bilingualism and arithmetic problem solving in Hispanic first graders. *Elementary School Journal*, *92*, 213 – 231.

Silver, R., Measelle, J., Armstrong, J., & Essex, M. (2005). Trajectories of classroom externalizing behavior: Contributions of child characteristics, family characteristics, and the teacher – child relationship during the school transition. *Journal of School Psychology, 43*, 39 – 60.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis*. New York: Oxford University Press.

Skinner, E., & Belmont, M. (1993). Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. *Journal of Educational Psychology, 85*, 571 – 581.

Spira, E., & Fischel, J. (2005). The impact of preschool inattention, hyperactivity, and impulsivity on social and academic development: A review. *Journal of Child Psychology and Psychiatry, 46*, 755 – 773.

Velting, O., & Whitehurst, G. (1997). Inattention-hyperactivity and reading achievement in children from low-income families: A longitudinal model. *Journal of Abnormal Child Psychology, 25*, 321–331.

Wentzel, K. (1997). Student motivation in middle school: The role of perceived pedagogical caring. *Journal of Educational Psychology, 89*, 411–419.

Williams, S., & McGee, R. (1994). Reading attainment and juvenile delinquency. *Journal of Child Psychology & Psychiatry & Allied Disciplines, 35*, 441–461.

Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock–Johnson Psychoeducational Battery–Revised*. Chicago: Riverside.

Child Development, November/December 2008, Volume 79, Number 6, Pages 1736–1751

Childhood Maltreatment and the Development of Relational and Physical Aggression: The Importance of a Gender-Informed Approach

Crystal Cullerton-Sen
University of Wisconsin–Stout

Adam R. Cassidy
University of Minnesota

Dianna Murray-Close
University of Vermont

Dante Cicchetti and Nicki R. Crick
University of Minnesota

Fred A. Rogosch
University of Rochester

This investigation examined the associations between maltreatment and aggression using a gender-informed approach. Peer ratings, peer nominations, and counselor reports of aggression were collected on 211 maltreated and 199 nonmaltreated inner-city youth (*M* age = 9.9 years) during a summer day camp. Maltreatment was associated with aggressive conduct; however, these effects were qualified by gender, maltreatment subtype, and the form of aggression under investigation. Findings revealed that maltreatment was associated with physical aggression for boys and relational aggression for girls. Physical abuse was associated with physically aggressive behaviors, but sexual abuse predicted relational aggression for girls only. Findings suggest that investigating the interaction between familial risk and gender is important in understanding aggressive behaviors of boys and girls.

A robust finding in the maltreatment literature is the association between childhood abuse and externalizing difficulties, particularly children's aggression toward peers (Klimes-Dougan & Kistner, 1990; Salzinger, Feldman, Hammer, & Rosario, 1993; Shields & Cicchetti, 2001; Sternberg, Baradaran, Abbott, Lamb, & Guterman, 2006). Unfortunately, our understanding of these links for boys and girls is limited in several important ways. First, work is needed that extends definitions of aggression to forms that are salient to girls as well as boys (Crick, 1995). Second, few studies have examined the moderating effects of gender on the association between maltreatment and gender-salient forms of aggression. Finally, there is a paucity of research on the ways in which subtypes of maltreatment may interact with gender when predicting aggression among boys and girls. The present investigation represents an initial attempt to elucidate these processes by examining the patterns of associations between maltreatment and aggression using a gender-informed developmental approach.

Maltreatment and Aggression

The first goal of the present study was to gain a gender-balanced understanding of the association between maltreatment and aggressive conduct. A number of mechanisms have been posited to underlie the relation between maltreatment and aggression. Due to the inconsistent care and relational toxicity that exist in maltreating families, maltreatment may disrupt children's internal working models (i.e., the view of self as worthy and competent and of others as caring and responsive), which in turn place them at risk for heightened aggression (Cicchetti & Toth, 2005; Rogosch, Cicchetti, & Aber, 1995; Shields, Ryan, & Cicchetti, 2001). Consistent with this perspective, Shields et al. (2001) found that negative and constrictive representations of caregivers significantly increased the likelihood that maltreated children would exhibit aggression with peers. Children who carry these internal representations forward into their relationships with peers are likely to readily perceive threats from peers, misinterpret social cues from

---

This research was supported, in part, by grants awarded to Dante Cicchetti, PhD, from the National Institute of Drug Abuse (DA 17741) and the Office of Child Abuse and Neglect and the Spunk Fund Inc. and to Nicki R. Crick, PhD, from the National Institute on Child Health and Development. We are grateful to the children, families, counselors, and research staff at the Mount Hope Family Center, Rochester, NY, who participated in this work.

Correspondence concerning this article should be addressed to Crystal Cullerton-Sen, School of Education, University of Wisconsin–Stout, 411 McCalmont Hall, Menomonie, WI 54751. Electronic mail may be sent to crystals@umn.edu.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7906-0014

others, and respond to peers with anger and aggression (e.g., Teisl & Cicchetti, 2008; Troy & Sroufe, 1987).

Findings from the maltreatment literature are also commensurate with a social-cognitive perspective of aggression (e.g., Bandura, 1986; Crick & Dodge, 1994). Physically abused children not only generate more aggressive solutions in the face of conflict but also view aggression as an effective way to meet their social and instrumental goals (Dodge, Bates, & Pettit, 1990; Herrenkohl, Egolf, & Herrenkohl, 1997; Shields & Cicchetti, 2001; Weiss, Dodge, Bates, & Pettit, 1992). In addition, children exposed to abuse and neglect encounter difficulties accurately interpreting social cues, with a greater tendency to perceive hostility and anger compared to children who have not been maltreated (e.g., Pine et al., 2005; Pollak, Cicchetti, Hornung, & Reed, 2000).

Maltreatment may also contribute to aggressive behavior by interfering with children's emotion regulation abilities. Children who experience maltreatment at the hands of significant others may internalize the experienced hostility, fear, and rejection in ways that lead to difficulties managing frustration and anger, competencies that are necessary for the successful resolution of peer conflicts (e.g., Shields & Cicchetti, 2001). In fact, maltreated children exhibit heightened arousal in response to other's distress or interpersonal conflict and display more anger than their nonmaltreated peers (Cummings, Hennessy, Rabideau, & Cicchetti, 1994; Klimes-Dougan & Kistner, 1990; Maughan & Cicchetti, 2002). Such difficulties have been found to mediate the relation between maltreatment and poor social competence with peers (Shields, Cicchetti, & Ryan, 1994).

A Gender-Informed Approach to the Study of Maltreatment and Aggression

Despite considerable theoretical and empirical work linking maltreatment and aggression, at present, we know very little about the ways in which gender may interact with maltreatment experiences in predicting aggression among boys and girls. We argue that a gender-informed approach capable of simultaneously modeling gender-balanced predictors and outcomes is needed. First, evidence suggests that boys and girls may differ in the ways they experience aggression (Archer & Coyne, 2005; Paquette & Underwood, 1999). Second, different etiologic risks may predict different aggression trajectories for boys and girls (e.g., Crick & Zahn-Waxler, 2003; Moffit & Caspi, 2001; Silverthorn & Frick, 1999). Based on this work, it would be important to examine how different forms of maltreatment, for example, physical and sexual abuse, interact with gender when predicting gender-relevant forms of aggression.

Relational aggression. A gender-informed approach should include relational as well as physical aggression for two reasons. First, although both boys and girls use physical and relational aggression, there are important differences. Boys tend to use physical force or make physical threats more than girls, and girls, when aggressive, are more likely to use relational, indirect, or social aggression over physical means (Archer & Coyne, 2005; Brendgen et al., 2005; Card, Stucky, Sawalani, & Little, 2008; Giles & Heyman, 2005; Ostrov & Keating, 2004). In this article, relational aggression will refer to overtly or covertly socially manipulating behaviors used to harm relationships between two or more individuals. These behaviors include withdrawing friendships, excluding others, or sabotaging someone's relationships with others by spreading vicious rumors (Crick, 1995; for discussion of indirect/social aggression, see Archer & Coyne, 2005; Björkqvist, Lagerspetz, & Kaukiainen, 1992; Galen & Underwood, 1997).

Second, mounting evidence suggests that girls are more attuned to and affected by relational conflicts than boys. Relative to boys, girls perceive relationally aggressive acts as mean and hurtful and experience heightened emotional and physiological distress in response to such behaviors (Coyne, Archer, & Elsea, 2006; Galen & Underwood, 1997; Giles & Heyman, 2005). Furthermore, whereas girls are more likely than boys to view relationally aggressive acts as wrong, boys are more likely than girls to view physically aggressive acts as wrong (Goldstein, Tisak, & Boxer, 2002; Murray-Close, Crick, & Galotti, 2006).

Relational aggression and maltreatment. Through many of the same mechanisms already discussed (e.g., biased social-cognitions, dysregulated affect), maltreated children may extend the relational hostility and confusion experienced within their family system into subsequent peer interactions by exhibiting relational aggression. For example, from a social information processing (SIP) perspective (Crick & Dodge, 1994), relationally aggressive children readily perceive threats in relational conflicts and use relationally aggressive behaviors to increase feelings of inclusion, control social connections, and reduce feelings of jealousy (Crick & Werner, 1998; Grotpeter & Crick, 1996). Maltreated children may learn through their interactions with caregivers that love and affection are given sporadically, can be withdrawn, and are used against others such that when confronted with relational threats in the peer context, they exclude or socially manipulate their peers. Likewise, because maltreatment and relational aggression are associated with heightened dysregulation in response to interpersonal conflicts (Coyne et al., 2006; Crick, Grotpeter, & Bigbee, 2002; Murray-Close & Crick, 2007; Shields et al., 1994), maltreated children may react to such conflicts with relational aggression. Preliminary research suggests that maltreated children exhibit significantly higher levels of relational aggression than their nonmaltreated peers (Rogosch & Cicchetti, 2005). However, relational aggression was included as part of a larger composite variable used to assess precursors of borderline personality features and was not of primary interest. Thus, the first goal of the present study was to investigate the links between maltreatment and both physical and relational aggression. We expected that maltreatment would emerge as an important predictor of both physical and relational aggression.

Gender as a moderator. The second goal of the present study was to examine whether maltreatment was associated with different forms of aggression for girls and boys. Within the maltreatment literature, the role of gender in the association between maltreatment and aggression has been relatively understudied. Due to the challenges inherent in this work, researchers have included samples of only one gender, used gender as a main effect variable, or been unable to examine the potential moderating effects of gender due to small samples (e.g., Calverley, Fischer, & Ayoub, 1994; Kitzmann, Gaylord, Holt, & Kenny, 2003; MacKinnon-Lewis et al., 1994; Trickett, McBride-Chang, & Putnam, 1994; Wolfe, Crooks, Lee, McIntyre-Smith, & Jaffe, 2003). Even when gender has been examined as a potential moderator, few investigations have found significant interactions between gender and maltreatment in the prediction of externalizing difficulties (Dodge, Pettit, & Bates, 1994; Sternberg et al., 2006). However, there are at least two reasons to expect that gender would moderate the relation between maltreatment and aggression using a gender-informed approach.

First, children at risk for aggressive behaviors due to experiences of maltreatment are likely to engage in gender-typical conduct when aggressive (i.e., relational aggression for girls, physical aggression for boys). Particularly during the early and middle childhood periods, children play in gender-segregated groups and exhibit behaviors that are consistent with the social norms of their peers (Rose & Rudolf, 2006). Furthermore, girls, but not boys, exhibit relational aggression in response to physiological stress caused by interpersonal provocations (e.g., Murray-Close & Crick, 2007). Thus, maltreated children may be especially likely to employ aggressive behaviors typical of their same-sex peer groups.

Second, research suggests that boys and girls differ in the ways in which they make meaning of their experiences (e.g., Rose & Rudolf, 2006) and in the value they place on various aspects of their relationships (e.g., Buhrmester, 1990; Bukowski, Newcomb, & Hartup, 1996). This may also be the case with respect to experiences of abuse and/or neglect. For example, girls may be more likely than boys to focus on the relational aspects of the maltreating environment, such as love withdrawal or the betrayal of trust within the parent–child dyad. Compared to boys, girls are more likely to focus on and value intimacy, trust, and closeness in their relationships (Coyne et al., 2006; Crick et al., 2002; Rose & Rudolf, 2006; Zahn-Waxler, Crick, Shirtcliff, & Woods, 2006) and to report dissatisfaction, emotional distress, and relational aggression when their relational needs are threatened (Crick et al., 2002; Galen & Underwood, 1997; Werner & Nixon, 2005). Therefore, because a focus on relational stressors is associated with heightened involvement in relational aggression (Crick et al., 2002; Murray-Close & Crick, 2007), we expected that maltreatment would predict especially high levels of relational aggression for girls.

In contrast, maltreated boys may be at particularly high risk for involvement in physical aggression. Boys tend to emphasize physical and instrumental competence in their social interactions with others (Bukowski et al., 1996; Rose & Rudolf, 2006) and are more likely to use physical aggression during instrumental conflict situations (Crick & Werner, 1998; Grotpeter & Crick, 1996). As such, maltreated boys may be particularly vulnerable to aspects of the maltreating environment that threaten their own or their loved ones' physical security such that they readily perceive and respond to such threats with physical aggression (Cummings et al., 1994; Dodge et al., 1990).

In sum, experiences of maltreatment may lead to different behavioral problems in boys and girls. We hypothesized that the effects of maltreatment on physical and relational aggression would be contingent on the gender of the child, such that maltreatment would predict higher levels of physical aggression among boys and higher levels of relational aggression among girls.

Gender, maltreatment subtypes, and aggression. The final goal of this study was to better understand the associations between specific maltreatment subtypes and aggression using a gender-informed approach. We expected that subtypes of abuse would differentially predict physical and relational aggression. Specifically, we hypothesized that physical abuse would be associated with physical forms of aggression. Physical abuse has been shown to disrupt SIP and emotional regulation, which in turn have been associated with physical aggression among physically abused children (Dodge, Pettit, Bates, & Valente, 1995; Klimes-Dougan & Kistner, 1990; Rogosch et al., 1995; Teisl & Cicchetti, 2008). In contrast, although there is a paucity of research in this area, we expected sexual abuse to be associated with relational aggression because sexual abuse violates aspects of relationships, such as trust and intimacy, which are known to be associated with relationally aggressive behavior patterns (Crick et al., 2002; Howes, Cicchetti, Toth, & Rogosch, 2000).

Specific maltreatment subtypes may be especially important in differentially predicting boys' and girls' aggressive conduct. Findings from the Third National Incidence Study indicate that boys are more likely to sustain physical injuries as the result of maltreatment (Sedlak & Broadhurst, 1996). Furthermore, some research suggests that physical abuse may have more deleterious effects on boys than on girls (e.g., Jaffe, Wolfe, Wilson, & Zak, 1986) and that physically abused boys are more likely than girls to exhibit physical aggression (Cummings et al., 1994). Thus, we hypothesized that physical abuse would predict higher levels of physical aggression, particularly for boys.

In contrast, girls are three times more likely than boys to be sexually abused (Sedlak & Broadhurst, 1996), and some researchers have proposed that sexual abuse may be a particularly salient risk factor for girls (e.g., Feiring, Taska, & Lewis, 1999). Girls are not only more likely than boys to experience sexual abuse and, when aggressive, use relationally aggressive behaviors, but compared to boys, they are also more likely to focus on and value intimacy and closeness in their relationships, aspects that become toxic and confusing where sexual abuse occurs (e.g., Howes et al., 2000; Kendall-Tackett, Williams, & Finkelhor, 1993). When needs of relational security are threatened, girls are more likely than boys to show signs of distress and to respond with relational aggression. Furthermore, both sexually abused and relationally aggressive girls display negative self-views and feel more guilt and regret in relation to their interactions with others (e.g., Calverley et al., 1994; Feiring et al., 1999; Moretti, Holland, & McKay, 2001). For these reasons, we hypothesized that sexual abuse would be associated with higher levels of relational aggression, particularly for girls.

Method

Participants

A total of 410 children who attended a weeklong day camp research program for inner-city youth participated in the present study. Demographic characteristics for the sample are shown in Table 1. Participants included 211 (123 boys) maltreated and 199 (90 boys) nonmaltreated children ranging in age from 6 years 1 month to 12 years 11 months (average age was 9.93 years). Maltreated and nonmaltreated children did not differ significantly on age, $F(1, 408) = 0.87$, *ns*. The majority of children were of an ethnic minority (83.7%). Fifty-six percent of participants were Black, 16.3% were White, 11.2% were Latino/Hispanic (non-Black), 13.8% were biracial, 0.7% were American Indian, and 2.2% were of other ethnicities. Ethnicity was significantly associated with maltreatment, $\chi^2(5, n = 410) = 11.83, p < .05$. Follow-up analyses sequentially comparing each ethnic group

Table 1
Demographic Characteristics of Sample by Gender and Maltreatment Status

		Gender		Maltreatment status	
Variable	Whole sample $N = 410$ (100%)	Boys $n = 213$ (52%)	Girls $n = 197$ (48%)	Maltreated $n = 211$ (51.5%)	Nonmaltreated $n = 199$ (48.5%)
Age	9.93 (1.81)	9.85 (1.86)	10.00 (1.75)	9.97 (1.86)	9.88 (1.75)
White	67 (16.34)	28 (6.83)	39 (9.51)	40 (9.76)	27 (6.59)
Minority	343 (83.67)	185 (45.12)	158 (38.54)	171 (41.71)	172 (41.95)
Black	228 (55.61)	120 (29.27)	108 (26.34)	106 (25.85)	122 (29.76)
Latino/Hispanic	46 (11.22)	29 (7.07)	17 (4.15)	21 (5.12)	25 (6.10)
Biracial	57 (13.90)	31 (7.56)	26 (6.34)	39 (9.51)	18 (4.39)
American Indian	3 (0.73)	3 (0.73)	0 (0)	1 (0.24)	2 (0.49)
Other	9 (2.20)	2 (0.49)	7 (1.71)	4 (0.98)	5 (1.22)
Boys	213 (52)			123 (30.00)	90 (21.95)
Girls	197 (48)			88 (21.46)	109 (26.59)

Note. Values for age = mean age (*SD*) in years. All other values indicate frequencies (percentages of whole sample).

against the remainder of the sample revealed a higher than expected level of maltreatment among biracial participants, $\chi^2(1, n = 410) = 7.62, p < .01$, and a lower than expected level of maltreatment among African American participants, $\chi^2(1, n = 410) = 5.08, p < .05$.

Participants in both maltreatment and comparison groups were from low-income, demographically similar backgrounds based on Hollingshead's (1975) four-factor index of social status (accounting for gender, occupation, education, and family configuration). All children were living with their biological mothers, most of whom were single parents. Maltreated children were recruited based on a history of child maltreatment as reported through the Monroe County (New York) Department of Social Services (DSS). Due to the need to protect the rights of confidentiality, families were initially approached by a DSS staff member. This DSS liaison met with a random sample of families on the DSS rosters who had children aged 6–12 years. The liaison explained the project to the mother of a target-age child and asked if she was willing to participate. The DSS liaison was careful to stress the voluntary nature of the project to avoid feelings of coercion and assured the mother that her refusal to participate would in no way affect her or her family's status with DSS or access to related services. Once mothers expressed an interest, the DSS liaison asked the mother sign a release form indicating her willingness to be contacted by research staff. Following a similar procedure, nonmaltreated participants were recruited from a random sample of families who received Temporary Assistance to Needy Families. Prior to enrolling in the study, mothers of both maltreated and nonmaltreated children provided written consent to allow trained project staff to examine any existing DSS records.

Several safeguards were used to determine the maltreatment status of the maltreated and nonmaltreated comparison groups. First, trained staff (doctoral students from Clinical and Social Sciences and clinical psychologists) carefully examined family records at the local Child Protection and Preventive Services to confirm the presence or absence of abuse within the family. Second, mothers of nonmaltreated children were interviewed to screen for maltreatment (see Cicchetti, Toth, & Manly, 2003). All maltreated children in the present study had a documented history of family abuse with Child Protection and Preventive Services at the DSS resulting in services through the agency; in contrast, nonmaltreated children had no history of abuse in these records or indicators of maltreatment based on interviews.

Specific maltreatment experiences were coded from DSS records using the Maltreatment Classification System (MCS; Barnett, Manly, & Cicchetti, 1993). Previous studies indicate that the MCS is both reliable and valid in classifying maltreatment subtypes (e.g., Bolger & Patterson, 2001; Manly, 2005). Consistent with this research, adequate reliability was obtained for each maltreatment subtype (kappas for physical, sexual, and emotional abuse were 1.00 and .90 for neglect).

The presence of sexual, physical, emotional abuse, and/or neglect in each family was assessed using criteria outlined by Barnett et al. (1993) based on the MCS operational criteria for maltreatment subtypes. The primary emphasis of the MCS subtypes definitions is on the parental behaviors that potentially result in psychological or physical harm to the child. *Sexual abuse* involves any attempted or actual sexual contact between a child and a caregiver for the caregiver's satisfaction or financial benefit. *Physical abuse* is the infliction of physical injury on a child other than by accident, such as beatings that result in bruises, choking, or burning a child. *Emotional maltreatment* occurs when caregivers prevent children from obtaining needed psychological safety and security, acceptance, self-esteem, and age-appropriate autonomy (e.g., ridiculing the child, abandoning the child, exposing the child to violence). *Neglect* is the failure to provide for children's basic physical needs and supervision or moral–legal or educational neglect. Neglect may result in lack of adequate food, shelter, or clothing or cause harm or potential harm from lack of supervision. Child protection services (CPS) narratives could include information that pertained to several forms of maltreatment. In such cases, information was scrutinized to determine the ways in which the parent's behavior threatened the child. For example, parent alcohol use may have resulted in the inability to provide adequate food or shelter and would thus be considered neglect, or it could have resulted in physically violent behaviors toward the child while under the influence and therefore be considered physical abuse. The number of children who experienced each subtype based on these definitions is provided in Table 2.

Design and Procedure

Participants attended a weeklong summer day camp for inner-city children (for details of this setting, see Cicchetti & Manly, 1990). During the summer camp, children were assigned to groups of approximately eight ($M = 7.57$, $SD = 1.57$) same-sex and same-age peers. Half of the members in each group were maltreated and half were nonmaltreated. Each group was led by three camp counselors who were unaware of participants' maltreatment status or

Table 2
*Frequencies and Percentages of Maltreated Children by Maltreatment Subtype (*N *= 410)*

	Maltreatment subtype			
	Sexual abuse, *n* (%)	Physical abuse, *n* (%)	Neglect, *n* (%)	Emotional maltreatment, *n* (%)
Presence (1)	31 (7.6)	104 (25.4)	169 (41.2)	147 (35.9)
Absence (0)	379 (92.4)	306 (74.6)	241 (58.8)	263 (64.1)

the study hypotheses. During the summer camp, children participated in a number of recreational and research activities, including the completion of instruments assessing the social behaviors of the peers in their group. In addition, camp counselors completed measures assessing the children's behavior in their group following each camp session.

Measures

Given the limitations of many single measures of aggressive conduct, children's physically and relationally aggressive behaviors were assessed using a multi-informant approach.

Peer ratings of aggression. First, children completed ratings of their peers' physically and relationally aggressive behaviors. Children were provided with one item depicting a physically aggressive child ("starts fights, says mean things, pushes or hits others") and one item depicting a relationally aggressive child ("when s/he is mad at someone, refuses to play or talk to the person, will try to get others not to like the person, will spread rumors or talk behind the person's back") and asked to rate how true each item was of each peer in their group on a scale of 0 (*not true*) to 2 (*very true*). The rating scores received from each peer in the group were averaged to yield an overall peer rating of physical aggression and relational aggression for each child.

Peer nominations of aggression. For the second measure of aggression, participants completed a peer nomination procedure. Group participation rates ranged from 75% to 100% ($M = 95\%$, $SD = 7\%$). Each child was asked to nominate one child in their group who fit the description of a physically aggressive child (this child "starts fights. He/she says mean things to other kids, or pushes them, or hits them") and a relationally aggressive child (this child, "when he/she is mad at someone, will refuse to play or talk to the person, will try to get others not to like the person, will spread rumors or talk behind the person's back"). The nominations received for each item were summed, standardized within group, and then standardized across all camp participants to yield total relational aggression and total physical aggression peer nomination scores.

Counselor report of aggression. The third measure of children's aggression was an instrument developed in previous research with teachers (Crick, 1996) and adapted to be used by camp counselors (Shields & Cicchetti, 2001). Counselors were provided with statements about physically aggressive behaviors (four items; e.g., "This child hits or kicks peers"), relationally aggressive behaviors (five items; e.g., "When mad at a peer, this child ignores the peer or stops talking to the peer"), and prosocial behaviors (four items; positively toned filler items) and asked to rate how true each statement was of the target child on a scale from 1 (*never true*) to 5 (*almost always true*). Each participating child was rated by two of the three counselors for their group. Subscales were computed by summing across physical and relational aggression items, respectively. These subscales were then averaged across counselors, yielding overall physical and relational aggression scores. Evidence for the validity of this measure has been demonstrated in previous research (Crick, 1996; Shields & Cicchetti, 2001). Cronbach's alpha demonstrated high reliability of the measures for this sample, with $\alpha = .97$ for physical aggression and $\alpha = .90$ for relational aggression.

Composite aggression scores. Composite physical and relational aggression scores were computed for each participant based on his or her counselor report, peer nomination, and peer ratings of aggression. Relational aggression scores and physical aggression scores from each method were transformed to z scores, summed, and then standardized across camp participants to yield a composite relational aggression score and a composite physical aggression score, respectively. The internal consistencies of the composite aggression scales were acceptable, with $\alpha = .70$ for relational aggression and $\alpha = .83$ for physical aggression. In addition, confirmatory factor analyses (CFA) using Mplus were conducted to examine the latent structure of the composite aggression variables. Results of the CFA indicated good fit with counselor reports, peer nominations, and peer ratings loading onto physical and relational aggression, respectively, and correlated residuals across informant types (comparative fit index [CFI] = 1.0, standardized root mean squared residual [SRMR] = .01). Moreover, all indicators exhibited large and significant loadings on their factor (*M* loading = 0.73, range = 0.50 – 0.97). Finally, nested model comparisons indicated that the model with separate physical and relational aggression factors provided significantly better fit than the model with one latent aggression factor, $\chi^2(1) = 8.34, p < .01$. These findings provide support for the validity of the composite physical and relational aggression factors in the present sample.

Means and standard deviations for each aggression measure are listed in Table 3. Peer ratings, peer nominations, and counselor reports of physical and relational aggression were significantly and moderately correlated (see Table 4). In addition, composite aggression scores for physical and relational aggression were strongly and significantly related to peer ratings, nominations, and counselor reports, with *r*s ranging from .83 to .91 and .76 to .82, respectively. Composite aggression scores were used in all analyses.

Table 4

Correlations Among Relational and Physical Aggression Measures (N = 410)

Variable	Correlations 1	2	3	4	5	6
1. Relational aggression—peer ratings	—					
2. Relational aggression—peer nominations	.53	—				
3. Relational aggression—counselor reports	.43	.37	—			
4. Physical aggression—peer ratings	.79	.59	.45	—		
5. Physical aggression—peer nominations	.54	.71	.35	.65	—	
6. Physical aggression—counselor reports	.60	.44	.50	.72	.51	—

Note. All correlations are statistically significant at $p < .001$.

Results

Descriptive information and correlations among age, race, gender, maltreatment, and aggression variables are presented in Table 5.

Maltreatment Status and Aggression

The first two goals of this study were to assess whether maltreatment status was associated with physical and relational aggression and to examine whether this association differed for boys and girls. We predicted that maltreatment would be associated with both physical and relational aggression. Furthermore, we expected that this relation would be moderated by gender, such that maltreatment would be more strongly related to physical aggression for boys and relational aggression for girls. Correlations between maltreatment status and aggression indicated that experiencing maltreatment was positively correlated with physical (phi = .19, $p < .001$) and relational (phi = .17, $p < .001$) aggression. To examine the role of gender in these relations, a series of

Table 3

Means and Standard Deviations of Relational and Physical Aggression Measures by Gender and Maltreatment Status

Measure	Whole sample *M* (*SD*)	Gender: Boys *M* (*SD*)	Gender: Girls *M* (*SD*)	Maltreatment status: Maltreated *M* (*SD*)	Maltreatment status: Nonmaltreated *M* (*SD*)
Relational aggression					
Peer ratings	0.01 (1.01)	0.17 (1.00)	−0.17 (0.98)	0.22 (1.10)	−0.22 (0.85)
Peer nominations	0.01 (1.01)	0.03 (1.02)	0.00 (1.00)	0.10 (1.07)	−0.07 (0.93)
Counselor reports	0.01 (1.01)	0.05 (0.98)	−0.04 (1.04)	0.12 (1.03)	−0.11 (0.98)
Composite variable	0.00 (1.00)	0.09 (0.97)	−0.10 (1.02)	0.17 (1.04)	−0.18 (0.92)
Physical aggression					
Peer ratings	0.01 (1.01)	0.25 (1.07)	−0.25 (0.87)	0.19 (1.09)	−0.18 (0.87)
Peer nominations	0.01 (1.01)	0.02 (1.02)	0.00 (1.00)	0.13 (1.09)	−0.11 (0.90)
Counselor reports	0.00 (1.01)	0.35 (1.15)	−0.37 (0.66)	0.21 (1.11)	−0.21 (0.85)
Composite variable	0.00 (1.00)	0.23 (1.10)	−0.24 (0.82)	0.19 (1.10)	−0.20 (0.84)

Note. See Table 1 for cell sizes.

Table 5

*Means, Standard Deviations, and Correlations Among Age, Ethnicity, Gender, Maltreatment, and Aggression (*N *= 410)*

	Correlations											
Variable	1	2	3	4	5	6	7	8	9	10	11	12
1. Age	—											
2. Ethnicity (Black)	.09†	—										
3. Ethnicity (White)	−.04	−.50***	—									
4. Ethnicity (biracial)	−.12*	−.45***	−.18***	—								
5. Gender	.04	−.02	.09†	−.02	—							
6. Maltreatment status	.03	−.11*	.07	.14**	−.13**	—						
7. Emotional maltreatment	.05	−.11*	.10†	.08†	−.15**	.73***	—					
8. Neglect	.07	−.04	.03	.08	−.07	.81***	.62***	—				
9. Physical abuse	.11*	−.07	.00	.09†	−.12*	.57***	.43***	.41***	—			
10. Sexual abuse	.04	−.10*	.12*	.05	.02	.28***	.25***	.25***	.22***	—		
11. Relational aggression	−.06	.11*	−.11*	.13*	−.10†	.17***	.11*	.15**	.16**	.06	—	
12. Physical aggression	−.06	.18***	−.13*	.03	−.24***	.19***	.14**	.16**	.22***	.02	.81***	—
M	9.93	0.56	0.16	0.14	0.48	0.51	0.36	0.41	0.25	0.08	0.00	0.00
SD	1.81	0.50	0.37	0.35	0.50	0.50	0.48	0.49	0.44	0.27	1.00	1.00

Note. Ethnicity (Black): 0 = *non-Black*, 1 = *Black*; ethnicity (White): 0 = *non-White*, 1 = *White*; ethnicity (biracial): 0 = *nonbiracial*, 1 = *biracial*; gender: 0 = *male*, 1 = *female*; maltreatment status: 0 = *nonmaltreated*, 1 = *maltreated*; emotional maltreatment: 0 = *nonemotionally maltreated*, 1 = *emotionally maltreated*; neglect: 0 = *nonneglected*, 1 = *neglected*; physical abuse: 0 = *nonphysically abused*, 1 = *physically abused*; and sexual abuse: 0 = *nonsexually abused*, 1 = *sexually abused*.

$^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

analyses of covariance (ANCOVAs) were conducted. Given the overlap between physical and relational aggression ($r = .81$, $p < .001$, in the present sample), the nonfocal form of aggression was entered as a covariate in all analyses. Age and ethnicity were also entered as covariates to control for the potential effects on aggression (Archer & Coyne, 2005; Cillessen & Mayeux, 2004). For these analyses, a four-category ethnicity variable was created representing Black, White, biracial, and other groups.

First, a 2 (maltreatment status: nonmaltreated vs. maltreated) × 2 (gender) ANCOVA was conducted with physical aggression serving as the dependent variable and age, ethnicity, and relational aggression entered as the covariates. As depicted in Table 6, the results indicated significant effects of ethnicity and relational aggression in predicting children's physically aggressive behaviors. Black children ($M = 0.16$, $SD = 1.05$) were significantly more likely to exhibit physical aggression than White children ($M = -0.28$, $SD = 0.72$), $p < .01$, and children of other ethnicities ($M = -0.38$, $SD = 0.78$), $p < .01$. Black and biracial children ($M = 0.08$, $SD = 1.12$) did not differ significantly on physical aggression. In addition, the effect for gender was significant, with boys exhibiting greater levels of physical aggression ($M = 0.14$, $SD = 1.10$) than girls ($M = -0.17$, $SD = 0.82$) based on the estimated marginal means. Finally, the Gender × Maltreatment Status interaction was significant. Follow-up planned comparisons revealed that maltreated boys ($M = 0.24$, $SD = 1.15$) were more physically aggressive than nonmaltreated boys ($M = 0.04$, $SD = 0.98$), $p < .05$; nonmaltreated girls ($M = -0.16$, $SD = 0.67$), $p < .001$; and maltreated girls ($M = -0.19$, $SD = 0.95$), $p < .001$.

Second, a 2 (maltreatment status) × 2 (gender) ANCOVA was conducted with relational aggression serving as the dependent variable and age, ethnicity, and physical aggression entered as the covariates. The results, shown in Table 7, indicated that physical aggression was a significant predictor of relational aggression. In addition, the effect for gender was significant, with girls displaying greater levels of relational aggression ($M = 0.11$, $SD = 1.02$) than boys ($M = -0.09$, $SD = 0.97$) based on the estimated marginal means. Finally, the Gender × Maltreatment Status interaction was significant. Planned comparisons indicated that maltreated girls ($M = 0.20$, $SD = 1.13$) were more relationally aggressive than nonmaltreated girls ($M = 0.03$, $SD = 0.89$), $p < .05$; maltreated boys ($M = -0.13$, $SD = 0.98$), $p < .01$; and nonmaltreated boys ($M = -0.05$, $SD = 0.94$), $p < .001$.

Subtypes of Maltreatment and Aggression

The third goal of the present study was to examine the association between subtypes of maltreatment and aggression and to investigate whether this relation differed for boys and girls. We expected that physical abuse would be associated with physical aggression, particularly for boys. In addition, we expected that sexual abuse would predict relational aggression, particularly for girls.

To examine the associations between maltreatment subtypes, gender, and physical and relational aggression, a series of hierarchical multiple regression analyses were computed. For each regression, age, ethnicity, and the nonfocal form of aggression were entered into Step 1. Ethnicity was dummy coded into several dichotomous variables representing "Black" (0 = *non-Black*, 1 = *Black*), "White" (0 = *non-White*, 1 = *White*), and "biracial" (0 = *nonbiracial*, 1 = *biracial*) groups, with "other" serving as the reference group. Maltreatment subtypes were then entered into Step 2. This allowed us to examine the unique effects of each subtype while controlling for the potential effects of experiencing other forms of maltreatment. Maltreatment subtypes were dummy coded to assess the effects of each subtype (e.g., 0 = *not sexually abused*, 1 = *sexually abused*) on aggressive conduct. Finally, to

Table 6
Summary of Analysis of Covariance for Variables Predicting Physical Aggression (N = 410)

Variable	*df*	*F*	Partial η^2	*p*
Age	2	0.26	.001	.61
Ethnicity	2	8.69**	.02	.003
Relational aggression	2	720.76***	.64	<.001
Gender	2	30.54***	.07	<.001
Maltreatment status	2	2.34	.006	.13
Maltreatment Status × Gender	2	4.23*	.01	.04

*$p < .05$. **$p < .01$. ***$p < .001$.

Table 7
Summary of Analysis of Covariance for Variables Predicting Relational Aggression (N = 410)

Variable	*df*	*F*	Partial η^2	*p*
Age	2	0.21	.001	.65
Ethnicity	2	0.54	.001	.47
Physical aggression	2	720.76***	.64	<.001
Gender	2	11.77**	.03	.001
Maltreatment status	2	0.75	.002	.39
Maltreatment Status × Gender	2	4.68*	.01	.03

*$p < .05$. **$p < .01$. ***$p < .001$.

investigate the effects of gender and its potential moderating role in the relation between maltreatment subtypes and aggression, gender (0 = *male*, 1 = *female*) and each Subtype × Gender interaction term were entered into Step 3. Due to the high correlation between relational and physical aggression, multicollinearity diagnostics were examined. Tolerance levels (0.3–0.34), variance inflation factors (2.94–3.3), and the standard error of the regression coefficients (.03) were in acceptable ranges for all regression analyses. Follow-up hierarchical regression analyses were computed separately for boys and girls to break down significant or marginally significant interactions between subtype of abuse and gender. In these analyses, age, ethnicity, nonfocal form of aggression, and subtypes from nonsignificant interaction terms were entered into Step 1 and maltreatment subtypes from significant interactions were entered into Step 2.

Physical aggression. The results of the regression examining the role of maltreatment subtypes in children's physically aggressive behaviors are presented in Table 8. Thus, findings revealed that age, ethnicity, relational aggression, maltreatment subtypes, gender, and the interactions between maltreatment subtypes and gender accounted for a substantial amount of variance (70%) associated with physical aggression. As shown in Table 8, experiences of maltreatment significantly predicted children's involvement in physical aggression, even after controlling for age, ethnicity, and relational aggression. Tests of individual effects indicated that children who were physically abused had higher levels of physical aggression than those who were not physically abused. The addition of gender and the interaction terms in Step 3 resulted in a significant, albeit small (3%), increase in the model R^2 ($p < .001$). Tests of individual effects indicated that boys were more physically aggressive than girls. The Gender × Sexual Abuse interaction was marginally significant ($p = .08$). Due to the conservative nature of these analyses, exploratory follow-up hierarchical regressions were conducted separately for boys and girls. Findings revealed that sexual abuse was significantly associated with physical aggression for girls, such that girls who were sexually abused exhibited lower levels of physical aggression than nonsexually abused girls ($p < .01$).

Relational aggression. The results of the regression examining the role of maltreatment subtypes in children's relationally aggressive behaviors are presented in Table 9. The results indicated that age, ethnicity, physical aggression, maltreatment subtypes, gender, and the interaction between maltreatment subtypes and gender accounted for a substantial amount of variance (68%) in the prediction of relational aggression.

As shown in Table 9, gender and experiences of maltreatment significantly predicted children's involvement in relational aggression, even after controlling for age, ethnicity, and physical aggression. The addition of gender and the interaction terms in Step 3 resulted in a significant, although small (2%), increase in the model R^2 ($p < .01$). In addition, the Gender × Sexual Abuse interaction was significant. Follow-up hierarchical regressions were conducted separately for boys and girls. Findings revealed that, as expected, sexual abuse was significantly associated with relational aggression for girls ($p < .01$) but not for boys.

Discussion

This investigation brings together three areas important to the study of development and advances our knowledge in a relatively understudied area of research, namely, maltreatment, gender, and relational aggression. Our goal in designing this study was to examine the relations between maltreatment and aggression using a gender-informed approach. In our models, we examined not only the associations between children's experiences of maltreatment and aggression but also the relations between specific subtypes of maltreatment and forms of aggression that are salient to both boys and girls (Crick et al., 2006; Feiring et al., 1999). Unlike prior research on maltreatment and aggression (e.g., Dodge et al., 1994; MacKinnon-Lewis et al., 1994; Shields & Cicchetti, 2001), we were able to discern the unique effects of gender and subtypes within a developmental perspective by including gender-relevant independent and dependent variables while controlling for nonfocal forms of aggression in these models.

The first goal of this study was to examine the association between childhood maltreatment and peer aggression by investigating whether children's experiences of maltreatment would significantly predict both physical and relational forms of aggression. Although maltreatment was positively associated with greater levels of both forms of aggression, our findings indicated that we must examine the moderating role of gender in order to understand the associations between maltreatment and aggression. This is consistent with the second goal of this investigation. In contrast to previous studies (e.g., Sternberg et al., 2006), we found that gender moderated these relations. As expected, whereas maltreated boys exhibited higher levels of physical aggression than

Table 8

Summary of Hierarchical Regression Analysis for Variables Predicting Physical Aggression (N = 410)

	Step 1			Step 2			Step 3		
Variable	*B*	*SE B*	β	*B*	*SE B*	β	*B*	*SE B*	β
Age	−0.02	0.02	−.03	−0.02	0.02	−.04	−0.02	0.02	−.03
Ethnicity									
Black vs. other	0.13	0.09	.06	0.14	0.09	.07	0.15	0.09	.08†
White vs. other	−0.03	0.11	−.01	−0.02	0.11	−.01	0.02	0.10	.01
Biracial vs. other	−0.13	0.11	−.05	−0.14	0.11	−.05	−0.13	0.11	−.05
Relational aggression	0.80	0.03	.80***	0.78	0.03	.78***	0.78	0.03	.78***
Emotional maltreatment				0.08	0.08	.04	0.09	0.10	.04
Neglect				0.01	0.08	.00	−0.05	0.10	−.03
Physical abuse				0.22	0.08	.10**	0.28	0.10	.12**
Sexual abuse				−0.20	0.11	−.05	0.04	0.16	.01
Gender							−0.24	0.08	−.12**
Emotional Maltreatment × Gender							−0.15	0.16	−.05
Neglect × Gender							0.18	0.15	.07
Physical Abuse × Gender							−0.20	0.15	−.06
Sexual Abuse × Gender							−0.40	0.22	−.08†
R^2	.66			.67			.70		
F for change in R^2	154.82***			4.02**			6.96***		

Note. Ethnicity was represented as three dummy variables with "other" (i.e., Black, White, or biracial) serving as the reference group.
$^{\dagger}p < .08$. $^{**}p < .01$. $^{***}p < .001$.

Table 9

*Summary of Hierarchical Regression Analysis for Variables Predicting Relational Aggression (*N *= 410)*

Variable	Step 1			Step 2			Step 3		
	B	*SE B*	β	*B*	*SE B*	β	*B*	*SE B*	β
Age	0.00	0.02	.01	0.00	0.02	.01	0.00	0.02	.00
Ethnicity									
Black vs. other	0.08	0.09	.04	0.08	0.09	.04	0.07	0.09	.03
White vs. other	0.08	0.11	.03	0.06	0.11	.02	0.04	0.11	.02
Biracial vs. other	0.36	0.11	.13**	0.35	0.11	.12**	0.34	0.11	.12**
Physical aggression	0.80	0.03	.80***	0.80	0.03	.80***	0.83	0.03	.83***
Emotional maltreatment				−0.05	0.08	−.03	−0.07	0.10	−.04
Neglect				0.06	0.08	.03	0.09	0.10	.04
Physical abuse				−0.08	0.08	−.03	−0.13	0.10	−.06
Sexual abuse				0.20	0.12	.05	−0.07	0.16	−.02
Gender							0.12	0.08	.06
Emotional Maltreatment × Gender							0.11	0.17	.04
Neglect × Gender							−0.10	0.16	−.04
Physical Abuse × Gender							0.16	0.15	.05
Sexual Abuse × Gender							0.48	0.23	.09*
R^2	.66			.66			.68		
F for change in R^2	155.70***			1.08			3.67**		

Note. Ethnicity was represented as three dummy variables with "other" (i.e., Black, White, or biracial) serving as the reference group.
$^*p < .05$. $^{**}p < .01$. $^{***}p < .001$.

maltreated girls and nonmaltreated children, maltreated girls displayed higher levels of relational aggression than all other groups. The associations between maltreatment and aggression for boys and girls may be explained in part by children's tendency to use forms of aggression that are consistent with the norms of their peer cultures, particularly during middle childhood. For example, Rose and Rudolf (2006) argue that children are more likely to exhibit sex-linked outcomes of adjustment (in this case, physical aggression among boys and relational aggression among girls) to the degree that they interact with same-sex peers. This would be particularly true in the current study where most participants were in middle childhood and all interacted in groups of same-sex peers.

Moreover, the unique associations (i.e., significant interactions) between maltreatment and relational and physical aggression for boys and girls indicate that girls and boys not only differ in their expression of aggression but may also internalize the experience of maltreatment in different ways. It is possible that boys and girls attend to and incorporate different aspects of their relationships with caregivers in making meaning of their relationships with peers. For example, maltreated children often display attentional as well as attributional biases that affect how they respond to others, understand their relationships with caregivers, and avoid or attend to threatening stimuli (Cicchetti & Curtis, 2005; Klimes-Dougan & Kistner, 1990; MacKinnon-Lewis et al., 1994; Rogosch et al., 1995; Shields et al., 2001). Future research should explore the ways in which maltreated boys and girls attend to various aspects of the caregiving system and how specific patterns of individual differences in attentional and attributional processes may influence peer interactions.

The third goal of this study was to extend current knowledge about the influence of subtypes of maltreatment in understanding the developmental trajectories of physical and relational aggression for both boys and girls. To address this issue, we conducted several hierarchical multiple regressions that enabled us to first assess the unique contributions of maltreatment subtypes in the prediction of physical and relational aggression and then to examine the contribution of gender and its potential moderating role on these relations. Findings revealed that maltreatment subtypes contributed significantly to the prediction of physical aggression above and beyond that accounted for by age, ethnicity, and relational aggression. As expected, physical abuse emerged as a stronger predictor of children's physical aggression than the other maltreatment subtypes examined. These findings are consistent with the view that physical abuse may initiate a pathway to externalizing difficulties, particularly physical aggression (e.g., Dodge et al., 1994), whereas experiences of neglect or sexual abuse may foster different developmental outcomes (Bolger & Patterson, 2001; Kim & Cicchetti, 2006). It is also possible that in order to understand how other maltreatment subtypes, such as sexual abuse, affect children's physical aggression, moderating variables, such as gender, need to be examined. Indeed, the findings revealed that the inclusion of gender and the Gender × Maltreatment Subtype interaction terms added a small but significant amount of variance in the prediction of physical aggression. Follow-up analyses revealed that sexual abuse was associated with lower levels of physical aggression for girls only.

Consistent with this view, results also indicated that gender and its moderating role must be considered when examining the associations between maltreatment subtypes and relational aggression. As hypothesized, findings revealed that sexual abuse was associated with higher levels of relational aggression for girls only, even after statistically controlling for the effects of age, ethnicity, and physical aggression. In contrast, without gender as a moderator, maltreatment subtypes did not contribute to the prediction of relational aggression.

Taken together, the findings are consistent with a gender-informed model of development and developmental psychopathology (Crick & Zahn-Waxler, 2003; Moffit & Caspi, 2001; Silverthorn & Frick, 1999; Zahn-Waxler et al., 2006) and suggest that we must consider specific maltreatment experiences, especially physical and sexual abuse, within the context of gender when predicting physical and relational aggression.

This investigation addressed several shortcomings present within the maltreatment and aggression literatures. First, unlike many prior studies (Kitzmann et al., 2003; Trickett et al., 1994; Wolfe et al., 2003), this investigation included a large sample of maltreated and nonmaltreated children, composed of both boys and girls, so that the moderating role of gender in the association between maltreatment subtype and aggression could be assessed. In contrast to other studies using gender as a control variable (e.g., Bolger & Patterson, 2000), we specifically examined the moderating effects of gender on the relation between maltreatment and aggression. Furthermore, the sample included demographically similar maltreatment and comparison groups, which controlled for the potentially confounding effects of environmental stressors on children's aggression, such as the risks associated with living in poverty. Finally, by investigating children's use of relational aggression among socioeconomically disadvantaged children, our findings add to the emergent literature on children's use of relational aggression among low-socioeconomic-status samples (Bonica, Arnold, Fisher, Zeljo, & Yershova, 2003; Sullivan, Farrell, & Kliewer, 2006) and call into question the notion often perpetuated among the popular press that relational aggression is predominately a problem of middle- and upper-class children (for a review, see Crick, 2007).

Implications, Limitations, and Future Directions

The current findings have several implications for future research. First, research is needed to understand the mechanisms through which maltreatment and its various subtypes affect the development of physical and relational aggression among boys and girls. Future research should explore whether boys and girls attend to and incorporate particular aspects of maltreating relational systems in ways that lead to their reliance on relational and physical aggression with peers. Of particular investigative interest would be how sexually and physically maltreated girls and boys allocate and control (e.g., initiate, inhibit, disengage) their attention to real and perceived threats, particularly emotional cues from the environment. For example, attentional biases have been found among physically abused children, such that they overattend to, more readily detect, and demonstrate heightened psychophysiological reactivity to expressions of anger (e.g., Pollak & Sinha, 2002). It is possible that boys and girls may exhibit different patterns of attention and control to different features of emotional expressions. For example, sexually abused girls have difficulties controlling intrusive thoughts and report heightened hyperarousal relative to boys (Feiring et al., 1999). These gender differences may have important implications for the development of physical and relational aggression among maltreated boys and girls.

Another important avenue for future research is to examine the ways in which attentional and attributional processes affect the meanings children make of their abusive experiences and how these interpretations relate to their use of aggression. Such work will be especially important for understanding the links between sexual abuse and relational aggression among girls. For example, compared to nonmaltreated girls, sexually abused adolescent girls have more negative, constrictive, and dissociative descriptions of themselves in response to the emotional tone set by those with whom they interact (e.g., Calverley et al., 1994). Emerging evidence also suggests that negative self-perceptions are related to relationally aggressive behaviors among girls but not among boys (Moretti et al., 2001). This finding may be associated with girls' tendency to feel guilt, regret, and loss of control following relationally aggressive events (Archer & Parker, 1994). Such feelings may begin during the sexually abusive experience, become internalized as distorted views of oneself in relation to others, and be expressed in the form of relational aggression, which in turn may bring about more guilt, regret, loss of control, and continued aggressive behaviors.

Given the disruption that maltreatment causes to children's ability to modulate their emotions and behaviors (Maughan & Cicchetti, 2002; Shields & Cicchetti, 2001), future work should also examine how emotional dysregulation relates to children's use of physical and relational aggression among maltreated and nonmaltreated children. Such difficulties have been shown to mediate the association between maltreatment and physical aggression (Shields et al., 2001), and given that relationally aggressive children have been shown to exhibit particularly heightened negative emotions and physiological reactivity in response to perceived relationship-oriented threats (Crick, 1995; Murray-Close & Crick, 2007), emotional dysregulation may serve as a vehicle through which maltreatment leads to children's use of relational aggression as well.

Finally, the current investigation represents a concurrent look at how relational and physical forms of aggression are related to maltreatment and how the development of these forms of aggression are highly contextualized (e.g., gender-specific) outcomes of an individual's interactions with his or her environments. However, replication research is needed. In addition, researchers should conduct longitudinal studies that explore the dynamic transactions within, as well as across, specific ecologies (e.g., family, peer, school) to better understand the links between maltreatment, aggression, and gender.

Conclusions

The findings from this study extend our current knowledge and have important implications for theory and research. This investigation represents the first study to specifically examine associations between maltreatment and relational aggression. Although the past decade has witnessed an increased focus on children's experience of relational aggression, few studies have investigated etiologic underpinnings of relational aggression or relevant familial risk factors associated with the development of relational aggression. This study adds to existing research by demonstrating that violent and relationally toxic maltreating family systems (often characterized by ambivalence, avoidance, enmeshment, and/or disorganization) affect not only the extent to which children exhibit physical aggression but also the degree to which they relationally manipulate their peers. This view is consistent with the second major contribution of this work. In this initial attempt to apply a gender-informed approach to the study of maltreatment and aggression, the findings suggest that we need to investigate the patterns of interactions between specific familial risk factors and aggression for both boys and girls. Indeed, our findings indicate that physical abuse may initiate a pathway to physically aggressive behaviors, whereas the developmental pathway for girls' aggression may be best understood when considering their sexual abuse histories and use of relational forms of aggression.

References

Archer, J., & Coyne, S. M. (2005). An integrated review of indirect, relational, and social aggression. *Personality and Social Psychology Review, 9*, 212–230.

Archer, J., & Parker, S. (1994). Social representations of aggression in children. *Aggressive Behavior, 20*, 101–114.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Upper Saddle River, NJ: Prentice Hall.

Barnett, D., Manly, J. T., & Cicchetti, D. (1993). Defining child maltreatment: The interface between policy and research. In D. Cicchetti & S. L. Toth (Eds.), *Advances in applied developmental psychology: Vol. 8. Child abuse, child development and social policy* (pp. 7–73). Norwood, NJ: Ablex.

Björkqvist, K., Lagerspetz, K., & Kaukiainen, A. (1992). Do girls manipulate and boys fight? Developmental trends in regard to direct and indirect aggression. *Aggressive Behavior, 18*, 117–127.

Bolger, K. E., & Patterson, C. J. (2000). Developmental pathways from child maltreatment to peer rejection. *Child Development, 72*, 549–568.

Bolger, K. E., & Patterson, C. J. (2001). Pathways from child maltreatment to internalizing problems: Perceptions of control as mediators and moderators. *Development and Psychopathology, 13*, 913–940.

Bonica, C., Arnold, D. H., Fisher, P. H., Zeljo, A., & Yershova, K. (2003). Relational aggression, relational victimization, and language development in preschoolers. *Social Development, 12*, 551–562.

Brendgen, M., Dionne, G., Girard, A., Boivin, M., Vitaro, F., & Perusse, D. (2005). Examining genetic and environmental effects on social aggression: A study of 6-year-old twins. *Child Development, 76*, 930–946.

Buhrmester, D. (1990). Intimacy of friendship, interpersonal competence, and adjustment during preadolescence and adolescence. *Child Development, 61*, 1101–1111.

Bukowski, W. M., Newcomb, A. F., Hartup, W. W. (Eds.). (1996). *The company they keep: Friendship in childhood and adolescence*. New York: Cambridge University Press.

Calverley, R. M., Fischer, K. W., & Ayoub, K. (1994). Complex splitting of self-representations in sexually abused adolescent girls. *Development and Psychopathology, 6*, 195–213.

Card, N. A., Stucky, B. D., Sawalani, G. M., & Little, T. D. (2008). Direct and indirect aggression during childhood and adolescence: A meta-analytic review of gender differences, intercorrelations, and relations to maladjustment. *Child Development, 79*, 1185–1229.

Cicchetti, D., & Curtis, W. J. (2005). An event-related potential study of the processing of affective facial expressions in young children who experienced maltreatment during the first year of life. *Development and Psychopathology, 17*, 641–677.

Cicchetti, D., & Manly, J. T. (1990). A personal perspective on conducting research with maltreating families: Problems and solutions. In E. Brody & I. Sigel (Eds.), *Family research: Vol. 2. Families at risk* (pp. 87–133). Hillsdale, NJ: Lawrence Erlbaum.

Cicchetti, D., & Toth, S. L. (2005). Child maltreatment. *Annual Review of Clinical Psychology, 1*, 409–438.

Cicchetti, D., Toth, S. L., & Manly, J. T. (2003). *Maternal child maltreatment interview.* Unpublished manuscript, Mount Hope Family Center, Rochester, NY.

Cillessen, A. H. N., & Mayeux, L. (2004). From censure to reinforcement: Developmental changes in the association between aggression and social status. *Child Development, 75*, 147–163.

Coyne, S., Archer, J., & Elsea, M. (2006). The frequency and harmfulness of indirect, relational, and social aggression. *Aggressive Behavior, 32*, 294–307.

Crick, N. R. (1995). Relational aggression: The role of intent attributions, feelings of distress, and provocation type. *Development and Psychopathology, 7*, 313–322.

Crick, N. R. (1996). The role of overt aggression, relational aggression, and prosocial behavior in the prediction of children's future social adjustment. *Child Development, 67*, 2317–2327.

Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information processing mechanisms in children's social adjustment. *Psychological Bulletin, 11*, 74–101.

Crick, N. R., Grotpeter, J. K., & Bigbee, M. A. (2002). Relationally and physically aggressive children's intent attributions and feelings of distress for relational and instrumental peer provocations. *Child Development, 73*, 1134–1142.

Crick, N. R., & Werner, N. E. (1998). Response decision processes in relational and overt aggression. *Child Development, 69*, 1630–1639.

Crick, N. R., & Zahn-Waxler, C. (2003). The development of psychopathology in females and males: Current progress and future challenges. *Development and Psychopathology, 15*, 719–742.

Cummings, E. M., Hennessy, K. D., Rabideau, G. J., & Cicchetti, D. (1994). Responses of physically abused boys to interadult anger involving their mothers. *Development and Psychopathology, 6*, 31–41.

Dodge, K. A., Bates, J. E., & Pettit, G. S. (1990). Mechanisms in the cycle of violence. *Science, 250*, 1678–1683.

Dodge, K. A., Pettit, G. S., & Bates, J. E. (1994). Effects of physical maltreatment on the development of peer relations. *Development and Psychopathology, 6*, 43–55.

Dodge, K. A., Pettit, G. S., Bates, J. E., & Valente, E. (1995). Social information-processing patterns partially mediate the effect of early physical abuse on later conduct problems. *Journal of Abnormal Psychology, 104*, 632–643.

Feiring, C., Taska, L., & Lewis, M. (1999). Age and gender differences in children's and adolescents' adaptation to sexual abuse. *Child Abuse & Neglect, 23*, 115–128.

Galen, B. R., & Underwood, M. K. (1997). A developmental investigation of social aggression among children. *Developmental Psychology, 33*, 589–600.

Giles, J. W., & Heyman, G. D. (2005). Young children's beliefs about the relationship between gender and aggressive behavior. *Child Development, 76*, 107–121.

Goldstein, S. E., Tisak, M. S., & Boxer, P. (2002). Preschoolers' normative and prescriptive judgments of relational and overt aggression. *Early Education and Development, 13*, 23–39.

Grotpeter, J. K., & Crick, N. R. (1996). Relational aggression, overt aggression and friendship. *Child Development, 67*, 2328–2338.

Herrenkohl, R. C., Egolf, B. P., & Herrenkohl, E. C. (1997). Preschool antecedents of adolescent assaultive behavior: A longitudinal study. *American Journal of Orthopsychiatry, 67*, 422–432.

Hollingshead, A. (1975). *Four-factor index of social status.* Unpublished manuscript, Yale University.

Howes, P. W., Cicchetti, D., Toth, S. L., & Rogosch, F. A. (2000). Affective, organizational, and relational characteristics of maltreating families: A systems perspective. *Journal of Family Psychology, 14*, 95–110.

Jaffe, P. G., Wolfe, D., Wilson, S., & Zak, L. (1986). Family violence and child adjustments: A comparative analysis of girls' and boys' behavioral symptoms. *American Journal of Psychiatry, 143*, 74–77.

Kendall-Tackett, K. A., Williams, L. M., & Finkelhor, D. (1993). Impact of sexual abuse on children: A review and synthesis of recent empirical studies. *Psychological Bulletin, 113*, 164–180.

Kim, J., & Cicchetti, D. (2006). Longitudinal trajectories of self-esteem processes and depressive symptoms among maltreated and nonmaltreated children. *Child Development, 77*, 624–639.

Kitzmann, K. M., Gaylord, N. K., Holt, A. R., & Kenny, E. D. (2003). Child witnesses to domestic violence: A meta-analytic review. *Journal of Consulting and Clinical Psychology, 71*, 339–352.

Klimes-Dougan, B., & Kistner, J. A. (1990). Physically abused preschoolers' responses to peer distress. *Developmental Psychology, 26*, 599–602.

MacKinnon-Lewis, C., Volling, B. L., Lamb, M. E., Dechman, K., Rabiner, D., & Curtner, M. E. (1994). A cross-contextual analysis of boys' social competence: From family to school. *Developmental Psychology, 30*, 325–333.

Manly, T. J. (2005). Advances in research definitions in child development. *Child Abuse & Neglect, 29*, 425–439.

Maughan, A., & Cicchetti, D. (2002). The impact of child maltreatment and interadult violence on children's emotion regulation abilities. *Child Development, 73*, 1525–1542.

Moffit, T. E., & Caspi, A. (2001). Childhood predictors that differentiate life-course persistent and adolescent-limited antisocial pathways among males and females. *Development and Psychopathology, 8*, 399–424.

Moretti, M. M., Holland, R., & McKay, S. (2001). Self-other representations and relational and overt aggression in adolescent girls and boys. *Behavioral Sciences and the Law, 19*, 109–126.

Murray-Close, D., & Crick, N. R. (2007). Gender differences in the association between cardiovascular reactivity and aggressive conduct. *International Journal of Psychophysiology, 65*, 103–113.

Murray-Close, D., Crick, N. R., & Galotti, K. M. (2006). Children's moral reasoning about physical and relational aggression. *Social Development, 15*, 345–372.

Ostrov, J. M., & Keating, C. F. (2004). Gender differences in preschool aggression during free play and structured interactions: An observational study. *Social Development, 13*, 255–277.

Paquette, J. A., & Underwood, M. K. (1999). Gender differences in young adolescents' experiences of peer victimization: Social and physical aggression. *Merrill-Palmer Quarterly, 45*, 242–266.

Pine, D. S., Mogg, K., Bradley, B. P., Montgomery, L., Monk, C. S., McClure, E., et al. (2005). Attention bias to threat in maltreated children: Implications for vulnerability to stress-related psychopathology. *American Journal of Psychiatry, 162*, 291–296.

Pollak, S. D., Cicchetti, D., Hornung, K., & Reed, A. (2000). Recognizing emotion in faces: Developmental effects of child abuse and neglect. *Developmental Psychology, 36*, 1025–1038.

Pollak, S. D., & Sinha, P. (2002). Effects of early experience on children's recognition of facial displays of emotion. *Developmental Psychology, 38*, 784–791.

Rogosch, F. A., & Cicchetti, D. (2005). Child maltreatment, attention networks, and potential precursors to borderline personality disorder. *Development and Psychopathology, 17*, 1071–1089.

Rogosch, F. A., Cicchetti, D., & Aber, J. L. (1995). The role of child maltreatment in early deviations in cognitive and affective processing abilities and later relationship problems. *Development and Psychopathology, 7*, 591–609.

Rose, A. R., & Rudolf, K. D. (2006). A review of sex differences in peer relationship processes: Potential trade-offs for the emotional and behavioral development of girls and boys. *Psychological Bulletin, 132*, 98–131.

Salzinger, S., Feldman, R. S., Hammer, M., & Rosario, M. (1993). The effects of physical abuse on children's social relationships. *Child Development, 64*, 169–187.

Sedlak, A. J., & Broadhurst, D. D. (1996). *Executive summary of the Third National Incidence Study of child abuse and neglect*. Retrieved October 27, 2007, from http://www.childwelfare.gov/pubs/statsinfo/nis3.cfm#foreward

Shields, A., & Cicchetti, D. (2001). Parental maltreatment and emotion dysregulation as risk factors for bullying and victimization in middle childhood. *Journal of Clinical Child Psychology, 30*, 349–363.

Shields, A., Cicchetti, D., & Ryan, R. M. (1994). The development of emotional and behavioral self regulation and social competence among maltreated school-age children. *Development and Psychopathology, 6*, 57–75.

Shields, A., Ryan, R. M., & Cicchetti, D. (2001). Narrative representations of caregivers and emotion dysregulation as predictors of maltreated children's rejection by peers. *Developmental Psychology, 37*, 321–337.

Silverthorn, P., & Frick, P. J. (1999). Developmental pathways to antisocial behavior: The delayed-onset pathway in girls. *Development and Psychopathology, 11*, 101–126.

Sternberg, K. J., Baradaran, L. P., Abbott, C. B., Lamb, M. E., & Guterman, E. (2006). Type of violence, age, and gender differences in the effects of family violence on children's behavior problems: A mega-analysis. *Developmental Review, 26*, 89–112.

Sullivan, T. N., Farrell, A. D., & Kliewer, W. (2006). Peer victimization in early adolescence: Association between physical and relational victimization and drug use, aggression, and delinquent behaviors among urban middle school students. *Development and Psychopathology, 18*, 119–137.

Teisl, M., & Cicchetti, D. (2008). Physical abuse, cognitive and emotional processes, and aggressive/disruptive behavior problems. *Social Development, 17*, 1–23.

Trickett, P. K., McBride-Chang, C., & Putnam, F. (1994). The classroom performance and behavior of sexually abused females. *Development and Psychopathology, 6*, 183–194.

Troy, M., & Sroufe, L. A. (1987). Victimization among preschoolers: Role of attachment relationship history. *Journal of the American Academy of Child and Adolescent Psychiatry, 26*, 166–172.

Weiss, B., Dodge, K. A., Bates, J. E., & Pettit, G. S. (1992). Some consequences of early harsh discipline: Child aggression and maladaptive social information processing style. *Child Development, 63*, 1321–1335.

Werner, N. E., & Nixon, C. L. (2005). Normative beliefs and relational aggression: An investigation of the cognitive bases of adolescent aggressive behavior. *Journal of Youth and Adolescence, 34*, 229–243.

Wolfe, D. A., Crooks, C. V., Lee, V., McIntyre-Smith, A., & Jaffe, P. G. (2003). The effects of children's exposure to domestic violence: A meta-analysis and critique. *Clinical Child and Family Psychology Review, 6*, 171–187.

Zahn-Waxler, C., Crick, N. R., Shirtcliff, R., & Woods, K. E. (2006). Gender and psychopathology: Conceptual, methodological, statistical, and substantive issues. In D. Cicchetti & D. Cohen (Eds.), *Handbook of developmental psychopathology: Vol. 3. Theory and method* (2nd ed., pp. 76–138). New York: Wiley.

Child Development, November/December 2008, Volume 79, Number 6, Pages 1752–1760

Neural Processing of Eye Gaze and Threat-Related Emotional Facial Expressions in Infancy

Stefanie Hoehl
Max Planck Institute for Human Cognitive and Brain Sciences

Tricia Striano
Hunter College

Combined with emotional expressions, eye gaze can provide essential information to indicate threat in the environment. The current study assessed the effects of eye gaze direction on infants' neural processing of fearful and angry faces. Event-related potentials were recorded from thirteen 7-month-old infants. Two face-sensitive posterior components, the N290 and P400, as well as a frontocentral negative component (Nc), indicating attentional arousal, were sensitive to eye gaze direction and emotion. A larger Nc was observed for angry faces with direct compared to averted eye gaze. Fearful faces elicited a larger N290 than angry faces, whereas angry faces elicited a more prominent P400 regardless of eye gaze direction. The findings are discussed in terms of early social cognitive and neural development.

Stimuli that indicate threat automatically capture attention and elicit fear (Öhmann, 2005). Attention is biased toward emotionally salient stimuli, especially emotional facial expressions, which possess a threat-related value, like angry and fearful faces (Vuilleumier, 2002). In adults, the neural processing of emotional facial expressions is affected by eye gaze direction, that is, whether eye gaze is directed at the observer or averted from the observer (Adams, Heather, Baird, Ambady, & Kleck, 2003; Sato, Yoshikawa, Kochiyama, & Matsumura, 2004). For instance, an angry face with eye gaze directed at the perceiver is more threatening than an angry face with eye gaze averted to the side. When faced with potential threat or when encountering an ambiguous situation, one essential reaction is to turn toward other people and evaluate their emotional expressions. Others' behavior and emotional expressions help to guide our own behavior, especially in ambiguous or potentially dangerous situations (e.g., Latané & Darley, 1970; Prentice & Miller, 1996). However, the developmental trajectory that leads to the highly efficient social threat detection system in human adults has only been investigated in parts.

Neonates already differentiate between direct and averted eye gaze and prefer to look at a face with the eyes directed toward them (Farroni, Csibra, Simion, & Johnson, 2002). By the age of 3 months, infants are able to follow other's eye gaze (D'Entremont, Hains, & Muir, 1997). Increasing empirical evidence suggests that young infants use the eye gaze of adults to shift their attention in the direction of socially relevant information (Hood, Willen, & Driver, 1998; Reid & Striano, 2005; Reid, Striano, Kaufman, & Johnson, 2004). Eye gaze detection is a very basic mechanism that can be observed very early in human ontogeny and that is also phylogenetically very old. It can be found even in nonmammalian species such as snakes and chickens (Emery, 2000). Primates use eye gaze to communicate with others, for instance, to express emotional states or to attract others' attention toward objects in their environment (Call, Hare, & Tomasello, 1995; Keating & Keating, 1982). It has been suggested that the ability to detect eye gaze direction has evolved in order to detect threat from potential predators (Haxby, Hoffmann, & Gobbini, 2002).

Infants are also sensitive to emotional expressions in face and voice from a very early age onward (see, e.g., de Haan & Nelson, 1998; Leppänen & Nelson, 2006; Walker-Andrews, 1997, for a review). For instance, by 6–7 months, infants discriminate most emotional expression pairings in habituation paradigms (de Haan & Nelson, 1998). Even though young infants possess remarkable skills regarding the discrimination and categorization of emotional facial

---

This research was supported by the Sofja Kovalevskaja Award granted by the Alexander von Humboldt Foundation and donated by the Federal Ministry of Education to T.S. We thank Christine Heinisch for assistance with testing. We are grateful to the infants and parents who participated and to the Universitätsfrauenklinik and the Klinikum St. Georg for support and assistance with recruitment.

Correspondence concerning this article should be addressed to Stefanie Hoehl, Max Planck Institute for Human Cognitive and Brain Sciences, Neuropsychology Department, Stephanstrasse 1a, 04103 Leipzig, Germany or Tricia Striano, Hunter College, 695 Park Avenue, New York, NY. Electronic mail may be sent to hoehl@cbs.mpg.de or tstriano@hunter.cuny.edu.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7906-0015

expressions, there is converging evidence that early social and emotional experience strongly affects the development of emotion processing (e.g., Parker, Nelson, & The Bucharest Early Intervention Project Core Group, 2005; Pollak, Cicchetti, Hornung, & Reed, 2000). As concluded by Leppänen and Nelson (2006), this indicates that the neural emotion processing system may not be innately specified but rather may become fine-tuned to the processing of emotional expressions with growing social experience.

Infants are especially susceptible to others' emotional feedback in ambiguous or disturbing situations. By 12 months, infants show social referencing; that is, they turn toward their caregivers and use referential emotional cues from caregivers to adjust their behavior accordingly (Moses, Baldwin, Rosicky, & Tidball, 2001; Sorce, Emde, Campos, & Klinnert, 1985). For instance, when placed on the edge of a visual cliff, that is, a Plexiglas surface covering an apparent depth, infants turn toward their caregivers to obtain information that disambiguates the apparently dangerous situation (see Feinman, Roberts, Hsieh, Sawyer, & Swanson, 1992, for a review). Facial and vocal expressions of emotion significantly affect infants' behavior when they are faced with a visual cliff (Sorce et al., 1985; Vaish & Striano, 2004). Infants cross the depth more often and also quicker when the mother directs joyful or interested expressions to them compared to angry or fearful expressions. In these studies, infants actively use emotional expressions provided by the caregiver to disambiguate the situation. However, it has to be noted that by using social referencing paradigms, it is often hard to distinguish between reactions to different emotions that share the same general value. For instance, infants may react in the same way to an angry and a fearful expression, both being negative in value, which makes it hard to draw conclusions on infants' discrimination between these expressions.

Despite the relevance of eye gaze detection and the perception of emotional facial expressions, little is known about how the developing human brain detects and responds to expressions of threat-related emotions in association with social referential cues. Studies on social referencing often require the infant to locomote or perform a certain action; therefore, mostly infants from the age of 12 months or older are tested (e.g., Sorce et al., 1985). This is why we know little about the emergence or precursors of social referencing in infants at younger ages.

In contrast to classic social referencing paradigms, event-related potential (ERP) measures do not require overt behavioral responses from the infant and may therefore be more sensitive measures of young infants' social cognitive skills (see also de Haan & Nelson, 1997). We conducted a visual ERP experiment to investigate the influence of eye gaze direction on the neural processing of negative emotional facial expressions. We tested 7-month-old infants because evidence from behavioral studies suggests that infants are able to discriminate and categorize angry and fearful facial expressions by that age (e.g., Schwartz, Izard, & Ansul, 1985; Serrano, Iglesias, & Leoches, 1992), though there is also an ERP study that did not show differential processing of fearful and angry facial expressions in 7-month-olds (Nelson & de Haan, 1996).

We focused our analyses on three components that have been most thoroughly investigated in infant face processing research: the N290, P400, and Nc. Both the early negative N290 component and the later occurring positive P400, component occur at posterior channel sites and are currently being discussed as potential precursors to the face-sensitive adult N170 (de Haan, Johnson, & Halit, 2003; Halit, Csibra, Volein, & Johnson, 2004). The N170 is related to structural encoding of physical information in faces (e.g., Bentin, Allison, Puce, Perez, & McCarthy, 1996) and is sensitive to the emotional expression in a face (Batty & Taylor, 2003). The adult N170 it is larger for emotionally significant compared to neutral faces and to biological motion compared to neutral control stimuli (e.g., Batty & Taylor, 2003; Puce & Perret, 2003). Functional similarities between the infant N290 and P400 on the one side and the adult N170 on the other side have been investigated in studies comparing the processing of faces to the processing of inverted faces, monkey faces, or visual noise (de Haan, Pascalis, & Johnson, 2002; Halit, de Haan, & Johnson, 2003; see de Haan et al., 2003, for a review).

In contrast to the adult N170, the infant N290 is also sensitive to eye gaze direction in 4-month-olds (Farroni et al., 2002). Specific responses of the N290 to different emotional facial expressions have not been found yet. However, Leppänen, Moulson, Vogel-Farley, and Nelson (2007) showed that the infant P400 is larger in response to fearful compared to happy or neutral facial expressions. Yet no study has investigated the effect of eye gaze direction in combination with fearful and angry facial expressions on the N290 and P400.

Another important component in the context of infant face processing is the midlatency negative central (Nc) component on frontocentral channels. The Nc has consistently been related to attentional orienting to salient stimuli (Courchesne, Ganz, & Norcia, 1981; Nelson, 1994), and its amplitude is closely associated with attention as measured by heart rate deceleration (Richards, 2003). In prior ERP studies, the Nc was sensitive to emotional expressions. For instance, Nelson and de Haan (1996) found an enhanced Nc in response to fearful compared to happy faces, indicating that 7-month-old infants allocated more attention toward fearful faces. The Nc was also found to discriminate between happy and angry faces (Grossmann, Striano, & Friederici, 2007) and between angry, neutral, and happy prosody (Grossmann, Striano, & Friederici, 2005).

Our intention was to investigate whether the face-sensitive components N290 and P400 as well as the Nc are affected by emotional facial expressions of fear and anger as a function of eye gaze direction.

Our prediction was that the more direct or obvious indicators of threat would elicit the higher attentional arousal and therefore the more prominent Nc on frontocentral channels. An angry face directing gaze at the perceiver is an immediate indicator of threat and potential aggression. A fearful face, in contrast, as an indicator of threat in the environment might be more threatening when averting gaze from the perceiver and thus potentially directing gaze at something threatening in the environment. Consistent with these considerations, behavioral studies with adults have shown that direct gaze facilitates processing of angry expressions and averted gaze facilitates processing of fearful expressions (Adams & Kleck, 2003). We therefore predict an enhanced Nc amplitude for angry faces with direct gaze and fearful faces with averted gaze compared to the other conditions.

We hypothesized that the P400 would be sensitive to emotional expression. In particular, we expected to find the larger responses to fearful compared to angry faces. In an ERP study by Batty and Taylor (2003) with adult subjects, the N170 was larger for fearful faces compared to the other basic emotions, including anger (see also Leppänen et al., 2007), though this finding could not be replicated with young children (Batty & Taylor, 2006).

Based on the findings by Farroni et al. (2002), we expected the N290 to be sensitive to gaze direction, with direct eye gaze eliciting the larger negative deflection. However, this hypothesis was only tentative because in the study by Farroni et al., a younger age group was tested and all faces were neutral.

Method

Participants

Participants were recruited from an existing database of families who had expressed interest in volunteering for research in infant development. Participants were White and of middle socioeconomic status. Thirteen infants (7 males and 6 females) were included in the final sample, with an average age of 7 months +/− 11 days. All infants were born full term (37 – 41 weeks) and were in the normal range for birth weight. Another 20 infants were tested but were excluded from the final sample as a result of fussiness ($n = 7$) or failing to reach the minimum requirements for adequate averaging of the ERP data ($n = 13$). This corresponds to the common dropout rate in infant ERP studies (e.g., Halit et al., 2004; Leppänen et al., 2007; Striano, Reid, & Hoehl, 2006). However, it has to be noted that the relatively high number of infants who failed to reach the minimum number of trials for averaging can likely be accounted for by the relatively high number of conditions and consequently required high number of stimulus presentations (see the Procedure subsection), which were very demanding for the infants in terms of sustaining attention and sitting still (see de Haan, 2007, on how the number of conditions are constrained in infant ERP research). To be included in the sample, an infant had to contribute at least 10 artifact-free trials per condition to their grand average, a criterion that is often used in infant ERP studies (Leppänen et al., 2007; Nelson & de Haan, 1996; Reid & Striano, 2005). For each of the conditions, the average number of trials included per infant was between 12.6 and 13.9 trials. The experiments were conducted with the understanding and the written consent of each participant's parent.

Stimuli

Stimulus material consisted of portrait photographs of one male and one female actor whose eyes were directed either to the front or to the left or to the right in a horizontal plane. In each of the four conditions, 50% of the pictures displayed each of the actors. In infant ERP studies most commonly only pictures from one or two different actors are presented to minimize the effects of differences in low-level psychophysiological properties between pictures on infants' neural responses. Original pictures were taken from the NimStim Face Stimulus Set (www.macbrain.org). Angry and fearful pictures were taken from both actors. Their irises were moved from the middle to the left and right corner of the eyes using Adobe Photoshop. Stimulus size was 17.5 × 22 cm.

Stimuli were rated by 20 adult subjects (12 females, mean age = 29.7 years). Participants were asked to identify the expressed emotion in a multiple-choice mode and to indicate to what degree the face signaled a direct threat for them on a 6-point scale. Expressed emotion was identified correctly in 93.8% of all trials for faces with direct eye gaze and 92.5% of all trials for faces with averted gaze, indicating that manipulation of the pupil did not affect recognition of emotional expression. Rating of threat as signaled by the face was mainly affected by a strong main effect of emotion, $F(19) = 23.07, p < .001$, indicating that angry faces were perceived as more threatening ($M = 2.96$, $SD = 0.31$) compared to fearful faces ($M = 1.41$, $SD = 0.16$) in general.

Procedure

Infants sat on their mother's lap in a dimly lit sound-attenuated and electrically shielded cabin at a viewing distance of 90 cm away from a 70 Hz 17-in. stimulus monitor. The experiment consisted of one block with 240 trials (containing 60 angry faces with direct eye gaze, 60 angry faces with averted eye gaze, 60 fearful faces with direct eye gaze, and 60 fearful faces with averted eye gaze). Stimulus presentation was controlled using the software ERTS (BeriSoft Corporation, Germany).

The four conditions were presented to the infant in a random order with the constraint that the same condition was not presented three times consecutively and that the number of presentations of each set of stimuli was balanced in every 16 trials. Each trial was preceded by a colorful attention attractor presented in the middle of the screen for 500 ms. Then, a stimulus picture was presented for 1,000 ms. After the presentation of every stimulus, the screen was blank for a random period of between 800 and 1,000 ms. The interstimulus interval (including the blank screen period and the fixation object) summed up to between 1,200 and 1,500 ms. If the infant became fussy or uninterested in the stimuli, the experimenter gave the infant a short break. During the break, the caregivers tried to reengage the infant if he or she got tired or bored or calm the infant if he or she got fussy. Sometimes caregiver and infant left the cabin for a couple of minutes. Then, the stimulus presentation and recording session were continued. The session ended when the infant's attention could no longer be attracted to the screen. Electroencephalogram (EEG) and electrooculogram (EOG) were recorded continuously, and the behavior of the infants was also video recorded throughout the session.

EEG Recording and Analyses

EEG was recorded continuously with Ag–AgCl electrodes from 19 scalp locations of the 10–20 system, referenced to the vertex (Cz), which were attached to a cap. Data were amplified via a Twente Medical Systems 32-channel REFA amplifier. Horizontal and vertical EOGs were recorded bipolarly. Sampling rate was set at 250 Hz. EEG data were rereferenced off-line to the mathematically combined linked mastoids. A band-pass filterwas set from 0.3 to 20 Hz.

The EEG recordings were segmented into epochs of waveform that comprised a 200-ms baseline featuring a colorful attention attractor and 1,000 ms of one static image featuring a face, as described earlier. For the elimination of electrical artifacts caused by eye and body movements, EEG data were rejected off-line whenever the standard deviation within a 200-ms gliding window exceeded 80 microvolts at EOG electrodes or 50 microvolts at any scalp electrode. Data were also visually edited off-line for artifacts and matched with the infant's recorded behavior. Only trials were included in which the infant had looked to the screen and displayed no eye movements on the video recording or on the EOG.

Initially, the grand average contained a very prominent early negative drift in one of the conditions (fearful–averted) starting directly after stimulus onset and lasting over the whole time course of the ERP on most of the channels. This drift could not be eliminated applying a baseline correction against mean voltage in the 200 ms period prior to stimulus onset. We thus decided to redo the analyses with a baseline correction against the mean voltage within the whole analyzed time window for all four conditions individually, which realigned the waveforms in a sufficient way, such that no unusual drifts were

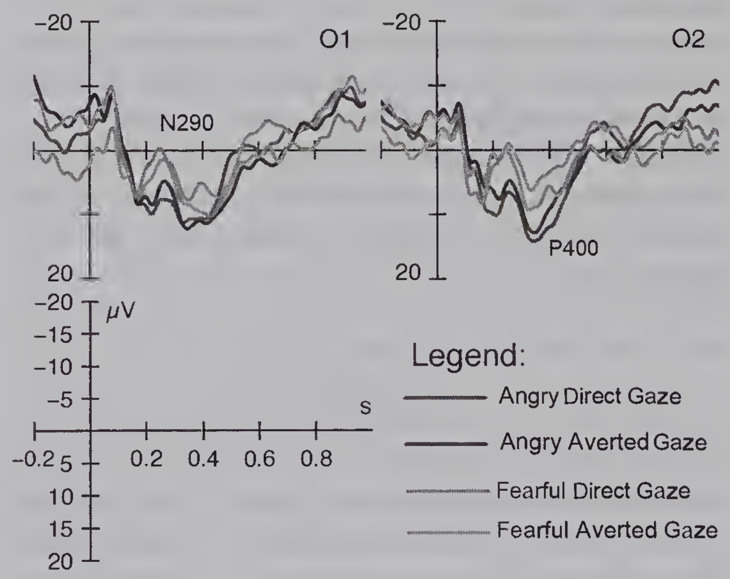


Figure 1. Grand average for all four conditions on occipital channels. Note that negative is plotted up in this figure. The N290 response is enhanced for fearful faces with direct and averted eye gaze (light gray lines) compared to angry faces with direct and averted eye gaze (dark gray and black lines). The P400 is more pronounced in conditions with angry facial expressions.

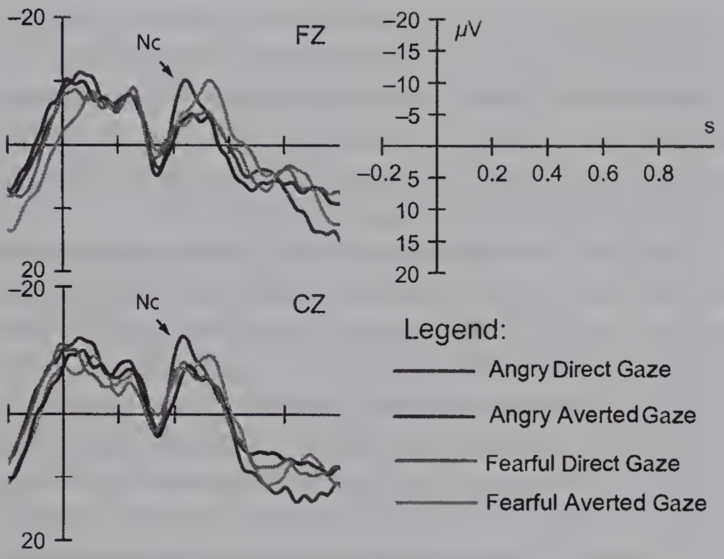


Figure 2. Grand average for all four conditions on frontocentral channels. Note that negative is plotted up in this figure. The Nc is larger for angry faces with direct eye gaze (dark gray lines) compared to angry faces with averted eye gaze (black lines). No similar effect was detected in the fearful face conditions.

evident in the final grand average (see Figures 1 and 2). Note that the same baseline correction was used for all four conditions.

For statistical analyses, three time windows were chosen to assess the effects of direct and averted eye gaze on infants' neural processing of angry and fearful faces. For the N290, minimal peak amplitude and latency were analyzed in an occipital channel group (O1 and O2) and on lateral parietal channels (P7 and P8) from 200 to 348 ms after stimulus onset. For the P400, maximal peak amplitude and latency were retrieved from 300 to 500 ms after stimulus onset on the same occipital and parietal channels. The Nc was investigated on a frontocentral channel group (FZ and CZ) from 348 to 600 ms after stimulus onset, assessing minimal peak amplitude and latency.

For all measures, a 2 × 2 repeated measures general linear model was performed with eye gaze (direct/averted) and emotion (angry/fearful) as independent factors.

Results

For the N290, a main effect of emotion was detected on occipital channels, $F(12) = 5.762$, $p = .03$. Minimal amplitude was smaller for fearful ($M = -5.85$ μV, $SD = 1.98$) compared to angry faces ($M = -1.49$ μV, $SD = 2.23$; see Figure 1). No other effects on amplitude or latency were found in this time range, and no effect was found on parietal channels.

For the P400, a main effect of emotion was found, $F(12) = 10.339$, $p = .007$. Maximal amplitude was larger for angry faces ($M = 21.10$ μV, $SD = 2.01$) than for fearful faces ($M = 15.37$ μV, $SD = 1.65$; see Figure 1). An additional peak-to-trough analysis was conducted to ensure that the effect found in this time window could not be accounted for by carryover effects due to the earlier N290 difference. Again, a significant main effect of emotion was found, $F(12) = 11.830$, $p = .005$. Maximal amplitude was larger for angry faces ($M = 19.64$ μV, $SD = 3.55$) than for fearful faces ($M = 9.52$ μV, $SD = 2.9$). No other effects on amplitude or latency were found in this time range, and no effect was found on parietal channels.

For the Nc, a significant Emotion × Eye Gaze interaction was found, $F(12) = 12.23$, $p = .004$. Post hoc t tests were applied to test our hypothesis that angry faces with direct eye gaze and fearful faces with averted eye gaze would elicit a larger Nc compared to the other conditions. In accordance to our expectations, infants reacted with an enhanced Nc to the angry faces with direct eye gaze ($M = -18.92$ μV, $SD = 7.12$) compared to angry faces with averted eye gaze ($M = -12.53$ μV, $SD = 5.11$), $t(12) = -3.62$, $p = .003$ (see Figure 2). For the fearful conditions, the Nc did not differ as a function of eye gaze direction. No other effects on amplitude or latency were found in this time range.

Additionally, we conducted statistic analyses on mean amplitude within the baseline on both channel groups because visual inspection of the ERPs indicated activity within the baseline, which can most probably be attributed to the colorful attention attractor that was presented prior to each stimulus and that was still presented during the baseline period. In particular, the negative deflection around stimulus onset on frontocentral channels may, according to its timing and topography, represent an Nc component in response to the central attractor. Though the attractor was the same for every stimulus picture, it may be that infants' reactions to the attractor were affected by the surrounding stimulus pictures. In other words, the presumably arousing angry and fearful face may have led to a more sensitive reaction and more pronounced Nc component in response to the attractor when compared to previous infant ERP studies. Variations in the baseline period on occipital channels may be related to the whole-window baseline correction that was used in the current study. A baseline correction to the mean voltage in the time window preceding stimulus onset usually realigns the waves in the baseline period on a similar level. This was not done in the present study. To still make sure that differences in later components are not affected by fluctuations in the prestimulus period, additional statistical analyses were conducted. Applying the same statistic analyses as described above, no significant effects on mean amplitude were found on frontocentral or occipital channels within the baseline period, indicating that the components that were explored in later time windows can be interpreted irrespective of fluctuations within the baseline.

Discussion

We conducted a visual ERP experiment with 7-month-old infants to investigate the effects of eye gaze and negative emotional expressions on the neural processing of unfamiliar faces. Differential processing of threat-related emotional expressions and eye gaze could be shown in several well-known components of the infant ERP.

When faces with angry expressions were presented with eye gaze directed at the infants attentional arousal was enhanced, as indicated by a larger Nc, compared to angry faces with averted eye gaze. No similar effect was found for fearful faces with direct or averted eye gaze. Our finding is in accordance with a previous ERP study in which 4-month-old infants showed enhanced neural processing of angry faces with direct eye gaze compared to averted eye gaze (Striano, Kopp, Grossmann, & Reid, 2006). An angry face that directs eye gaze at the perceiver is a clear signal for potential threat. Presumably, infants are already sensitive to the relevant social signals, which indicate threat in a social interaction even if only static two-dimensional images are presented. This highlights the social relevance and emotional valence of angry faces with direct eye contact already very early in ontogeny. However, alternative explanations should be taken into consideration. It may be that angry faces with direct eye gaze are merely less familiar to the infant compared to the other conditions. If that was the case, it may be that it was mainly the novelty of stimuli in this condition that elicited an enhanced Nc. It is known that mothers' expression of anger toward their infants does not increase until infants start to locomote (Campos et al., 2000). Experience and familiarity with emotional expressions are important factors that determine infants' discrimination and categorization abilities (Ludemann & Nelson, 1988) and probably affect the neural processing of emotional facial expressions (Parker et al., 2005). However, fearful faces with direct eye gaze should be no less familiar to the infants than angry faces. Still, the amplitude of the Nc was not enhanced for the fearful faces with direct eye gaze. Unfortunately, we have not tested infants' a priori familiarity with emotional facial expressions; therefore, a familiarity effect cannot be ruled out.

Importantly, our findings demonstrate that it was not direct eye gaze per se that elicited an enhanced Nc, as direct eye contact in association with a fearful expression did not have the same effect as in the context of an angry face.

In addition, differences in the face-sensitive posterior components N290 and P400 were assessed. In accordance with results from a previous ERP study with adults (Batty & Taylor, 2003), the N290 was significantly larger for fearful faces compared to angry faces, regardless of eye gaze direction. This main effect of emotion on amplitude of the N290 was evident on central occipital channels (O1 and O2) and not on more lateral channel sites (P7 and P8), as found in the study by Batty and Taylor (2003). As reviewed by de Haan et al. (2003), the infant N290 in general shows a more central distribution compared to the adult N170, which may account for the more centrally distributed emotion effect in the current study. There are also studies investigating the adult N170 that did not find effects of emotional expressions on N170 amplitude (see Eimer & Holmes, 2007, for a review). It may be that the infant N290 is more sensitive to emotional expressions in the face than the adult component. There is evidence that the N290 component, compared to the adult N170, is less fine-tuned to the structural encoding of faces (Johnson et al., 2005), which might be relatively independent from emotional expression processing in adults (Eimer & Holmes, 2007). For instance, the infant N290 is sensitive to gaze direction in static pictures, whereas the N170 in adults is not (Farroni et al., 2002). This lack of specificity of infants' neural responses has been seen as evidence for an interactive specialization of the infant brain toward the highly specialized adult social brain network during the first years of life (Johnson, 2000).

The P400 on central occipital channels was substantially more positive as a response to angry compared to fearful faces. As indicated by the additional peak-to-trough analysis, the latter finding is not driven by carryover effects due to the earlier occurring negative N290. Leppänen et al. (2007) found effects of emotional expression on P400 amplitude on more lateral semimedial posterior channels. However, in the study by Leppänen et al., a different configuration of channel sites was used which makes it difficult to directly compare between results from both studies. Further, Leppänen et al. did not test infants' reactions to angry versus fearful faces but compared happy, surprised, and angry faces. Most importantly, however, Leppänen et al. rereferenced their data to the average reference, which was not possible for us due to a smaller number of recording channels. The disadvantage of the linked mastoid reference as used in the current study is that ERPs on channel sites near the mastoids, namely P7 and P8, are attenuated, which makes it difficult to detect effects on these channels (Dien, 1998).

Both the N290 and the P400 are being discussed as potential precursors to the adult N170 as they share functional properties with the adult component and become more finely tuned to the processing of human faces during development in the 1st year (see de Haan et al., 2003, for a review). It has been proposed that the N290 and P400 reflect neural processes that later in development become integrated in a single component, the N170 (de Haan et al., 2003). Our findings suggest that in terms of emotional facial expression processing, the N290 may be more similar to the adult N170 than the P400. But it has to be kept in mind that such conclusions can only be drawn under the assumption that similar processes affect infant and adult face processing, which is rather unlikely. Presumably, familiarity with facial expressions is another factor that might affect even early neural responses. This assumption, however, remains to be tested.

In contrast to a previous study by Farroni et al. (2002), no main effect of eye gaze direction was found for either of the two posterior components. However, in the current study, a different age group was tested and emotionally significant instead of neutral stimuli were applied. Furthermore, methodological differences may account for the discrepancy. We rereferenced EEG data to the linked mastoids instead of the average of all applied channels, used a different baseline correction, and, due to our recording system, we were not able to assess the same channels that were investigated by Farroni et al.

The components investigated in the current study may be part of a developing neural threat detection system. A relationship between amygdala activation and the perception of threatening stimuli and fear is a very consistent finding in studies with adults (Adolphs, 2002; LeDoux, 2003). Empirical evidence even suggests a direct subcortical pathway from the thalamus, a relay station for sensory input, to the amygdala, without passing sensory cortices (Morris, Öhman, & Dolan, 1999). This provides a very fast and efficient way to drag attention in a bottom-up kind of fashion to potentially threatening information, before conscious and more elaborate processing of the fear-eliciting stimuli takes place. It may be that attention is drawn quickly and automatically to angry faces with direct eye gaze already in infancy as they are very immediate indicators of threat and elicit activation in subcortical structures, which then modulate activation in cortical structures related to attentional processes (see also Leppänen et al., 2007).

The current findings support evidence from behavioral studies in which infants at the age of 7 months used emotional expressions provided by adults in association with referential cues to disambiguate potentially dangerous situations or ambiguous behavior (Striano & Rochat, 1999; Striano & Vaish, 2005). According to a definition by Feinman et al. (1992), this behavior can be called social referencing. Feinman et al. defined three elements of social referencing in infancy: (a) seeking information in adults in novel or ambiguous situations, (b) associating the referent with the social message, and (c) regulating behavior in response to the provided information. Still, the developmental course or even onset of social referencing is currently unknown because almost all studies in this field investigated infants from the age of 12 months onward. Because infants are constantly encountering situations they might perceive as novel and ambiguous, it makes sense that they try to obtain information from other people in form of social cues, like eye gaze and facial expressions, much earlier.

In sum, the current findings further our knowledge of neural face processing in infancy. Amplitudes of two face-sensitive posterior components were affected by angry and fearful facial expressions. The Nc was also sensitive to our manipulation of eye gaze direction, indicating enhanced attentional arousal when angry faces were shown with eye gaze directed at the infants. We conclude that infants at the age of 7 months are already sensitive to social signals that indicate threat in the environment. This highly adaptive early competency is most likely an important precursor to social referencing behavior.

References

Adams, R. B., Heather, L. G., Baird, A. A., Ambady, N., & Kleck, R. E. (2003). Effects of gaze on amygdala sensitivity to anger and fear faces. *Science*, *300*, 1536.

Adams, R. B., & Kleck, R. E. (2003). Perceived gaze direction and the processing of facial displays of emotion. *Psychological Science*, *14*, 644–647.

Adolphs, R. (2002). Neural systems for recognizing emotion. *Current Opinion in Neurobiology*, *12*, 169–177.

Batty, M., & Taylor, M. J. (2003). Early processing of the six basic facial emotional expressions. *Cognitive Brain Research*, *17*, 613–620.

Batty, M., & Taylor, M. J. (2006). The development of emotional face processing during childhood. *Developmental Science*, *9*, 207–220.

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience*, *8*, 551–565.

Call, J., Hare, B. A., & Tomasello, M. (1995). Chimpanzee gaze following in an object-choice task. *Animal Cognition*, *1*, 89–99.

Campos, J. J., Anderson, D. I., Barbu-Roth, M. A., Hubbard, E. M., Hertenstein, M. J., & Witherington, D. (2000). Travel broadens the mind. *Infancy, 1*, 149 – 219.

Courchesne, E., Ganz, L., & Norcia, A. M. (1981). Event-related brain potentials to human faces in infants. *Child Development, 52*, 804 – 811.

de Haan, M. (2007). *Infant EEG and event-related potentials.* Hove, England: Psychology Press.

de Haan, M., Johnson, M. H., & Halit, H. (2003). Development of face-sensitive event-related potentials during infancy: A review. *International Journal of Psychophysiology, 51*, 45 – 58.

de Haan, M., & Nelson, C. (1997). Recognition of the mother's face by six-month-old infants: A neurobehavioral study. *Child Development, 68*, 187 – 210.

de Haan, M., & Nelson, C. A. (1998). Discrimination and categorization of facial expressions of emotion during infancy. In A. M. Slater (Ed.), *Perceptual development: Visual, auditory, and language perception in infancy* (pp. 287 – 309). London: University College London Press.

de Haan, M., Pascalis, O., & Johnson, M. (2002). Specialisation of neural mechanisms underlying face recognition in human infants. *Journal of Cognitive Neuroscience, 14*, 199 – 209.

D'Entremont, B., Hains, S. M. J., & Muir, D. W. (1997). A demonstration of gaze following in 3- to 6-month-olds. *Infant Behavior and Development, 20*, 569 – 572.

Dien, J. (1998). Issues in the application of the average reference: Review, critiques, and recommendations. *Behavior Research Methods, Instruments and Computers, 30*, 34 – 43.

Eimer, M., & Holmes, A. (2007). Event-related brain potential correlates of emotional face processing. *Neuropsychologia, 45*, 15 – 31.

Emery, N. J. (2000). The eyes have it: Neuroethology, function and evolution of social gaze. *Neuroscience and Biobehavioral Reviews, 24*, 581 – 604.

Farroni, T., Csibra, G., Simion, F., & Johnson, M. H. (2002). Eye contact detection in humans from birth. *Proceedings of the National Academy of Science, 99*, 9602 – 9605.

Feinman, S., Roberts, D., Hsieh, K.-F., Sawyer, D., & Swanson, D. (1992). A critical review of social referencing in infancy. In S. Feinman (Ed.), *Social referencing and the social construction of reality in infancy* (pp. 15 – 54). New York: Plenum.

Grossmann, T., Striano, T., & Friederici, A. D. (2005). Infants' electric brain responses to emotional prosody. *NeuroReport, 16*, 1825 – 1828.

Grossmann, T., Striano, T., & Friederici, A. D. (2007). Developmental changes in infants' processing of happy and angry facial expressions: A neurobehavioral study. *Brain and Cognition, 64*, 30 – 41.

Halit, H., Csibra, G., Volein, A., & Johnson, M. (2004). Face-sensitive cortical processing in early infancy. *Journal of Child Psychology and Psychiatry, 45*, 1228 – 1234.

Halit, H., de Haan, M., & Johnson, M. H. (2003). Cortical specialization for face processing: Face-sensitive event-related potential components in 3- and 12-month-old infants. *NeuroImage, 19*, 1180 – 1193.

Haxby, J. V., Hoffmann, E. A., & Gobbini, M. I. (2002). Human neural systems for face recognition and social communication. *Biological Psychiatry, 51*, 59 – 67.

Hood, B. M., Willen, J. D., & Driver, J. (1998). Adults eyes trigger shifts of visual attention in human infants. *Psychological Science, 9*, 131 – 134.

Johnson, M. H. (2000). Functional brain development in infants: Elements of an interactive specialization framework. *Child Development, 71*, 75 – 81.

Johnson, M. H., Griffin, R., Csibra, G., Halit, H., Farroni, T., de Haan, M., et al. (2005). The emergence of the social brain network: Evidence from typical and atypical development. *Development and Psychopathology, 17*, 599 – 619.

Keating, C. F., & Keating, E. G. (1982). Visual scan patterns of rhesus monkeys viewing faces. *Perception, 35*, 425 – 475.

Latané, B., & Darley, J. M. (1970). *The unresponsive bystander: Why doesn't he help?* New York: Appleton-Century-Crofts.

LeDoux, J. (2003). The emotional brain, fear and the amygdala. *Cellular and Molecular Neurobiology, 23*, 727 – 738.

Leppänen, J. M., Moulson, M. C., Vogel-Farley, V. K., & Nelson, C. A. (2007). An ERP study of emotional face processing in the adult and infant brain. *Child Development, 78*, 232 – 245.

Leppänen, J. M., & Nelson, C. A. (2006). The development and neural bases of recognizing of facial emotion. In R. Kail (Ed.), *Advances in child development and behavior* (pp. 207 – 246). Amsterdam: Elsevier.

Ludemann, P., & Nelson, C. A. (1988). The categorical representation of facial expressions by 7-month-old infants. *Developmental Psychology, 24*, 492 – 501.

Morris, J. S., Öhman, A., & Dolan, R. J. (1999). A subcortical pathway to the right amygdala mediating "unseen" fear. *Proceedings of the National Academy of Sciences U.S.A., 96*, 1680 – 1685.

Moses, L. J., Baldwin, D. A., Rosicky, J. G., & Tidball, G. (2001). Evidence for referential understanding in the emotions domain at twelve and eighteen months. *Child Development, 72*, 718 – 735.

Nelson, C. A. (1994). Neural correlates of recognition memory in the first postnatal year of life. In G. Dawson & K. Fischer (Eds.), *Human behavior and the developing brain* (pp. 269 – 313). New York: Guilford.

Nelson, C. A., & de Haan, M. (1996). Neural correlates of infants' visual responsiveness to facial expressions of emotion. *Developmental Psychobiology, 29*, 577 – 595.

Öhmann, A. (2005). The role of the amygdala in human fear: Automatic detection of threat. *Psychoneuroendocrin, 30*, 953 – 958.

Parker, S. W., Nelson, C. A., & The Bucharest Early Intervention Project Core Group. (2005). An event-related potential study of the impact of institutional rearing on face recognition. *Development and Psychopathology, 17*, 621 – 663.

Pollak, S. D., Cicchetti, D., Hornung, K., & Reed, A. (2000). Recognizing emotion in faces: Developmental effects of child abuse and neglect. *Developmental Psychology, 36,* 679–688.

Prentice, D. A., & Miller, D. T. (1996). Pluralistic ignorance and the perpetuation of social norms by unwitting actors. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 29, pp. 161–209). San Diego, CA: Academic Press.

Puce, A., & Perret, D. (2003). Electrophysiology and brain imaging of biological motion. *Philosophical Transactions of the Royal Society of London, 358,* 435–445.

Reid, V. M., & Striano, T. (2005). Adult gaze influences infant attention and object processing implications for cognitive neuroscience. *European Journal of Neuroscience, 21,* 1763–1766.

Reid, V. M., Striano, T., Kaufman, J., & Johnson, M. H. (2004). Eye gaze cueing facilitates neural processing of objects in 4-month-old infants. *NeuroReport, 15,* 2553–2555.

Richards, J. E. (2003). Attention affects the recognition of briefly presented visual stimuli in infants: An ERP study. *Developmental Science, 6,* 312–328.

Sato, W., Yoshikawa, S., Kochiyama, T., & Matsumura, M. (2004). The amygdala processes the emotional significance of facial expressions: An fMRI investigation using the interaction between expression and face direction. *NeuroImage, 22,* 1006–1013.

Schwartz, G., Izard, C., & Ansul, S. (1985). The 5-month-olds' ability to discriminate facial expressions of emotions. *Infant Behavior and Development, 8,* 65–77.

Serrano, J. M., Iglesias, J., & Leoches, A. (1992). Visual discrimination and recognition of facial expressions of anger, fear and surprise in four- to six-month-old infants. *Developmental Psychobiology, 25,* 411–425.

Sorce, J. F., Emde, R. N., Campos, J., & Klinnert, M. D. (1985). Maternal emotional signalling: Its effect on the visual cliff behavior of 1-year-olds. *Developmental Psychology, 21,* 195–200.

Striano, T., Kopp, F., Grossmann, T., & Reid, V. M. (2006). Eye contact influences neural processing of emotional expressions in 4-month-old infants. *Social Cognitive and Affective Neuroscience, 1,* 87–95.

Striano, T., Reid, V. M., & Hoehl, S. (2006). Neural mechanisms of joint attention in infancy. *European Journal of Neuroscience, 23,* 2819–2823.

Striano, T., & Rochat, P. (1999). Developmental link between dyadic and triadic social competence in infancy. *British Journal of Developmental Psychology, 17,* 551–562.

Striano, T., & Vaish, A. (2005). Seven- to 9-month-old infants use facial expressions to interpret others' actions. *British Journal of Developmental Psychology, 24,* 753–760.

Vaish, A., & Striano, T. (2004). Is visual reference necessary? Contributions of facial versus vocal cues in 12-month-olds' social referencing behavior. *Developmental Science, 7,* 261–269.

Vuilleumier, P. (2002). Facial expression and selective attention. *Current Opinion in Psychiatry, 15,* 291–300.

Walker-Andrews, A. S. (1997). Infants' perception of expressive behaviors: Differentiation of multimodal information. *Psychological Bulletin, 121,* 437–456.

Child Development, November/December 2008, Volume 79, Number 6, Pages 1761–1776

Fetal Substance Exposure and Cumulative Environmental Risk in an African American Cohort

Chie Yumoto, Sandra W. Jacobson, and Joseph L. Jacobson
Wayne State University

Two models of vulnerability to socioenvironmental risk were examined in 337 African American children (M = 7.8 years) recruited to overrepresent prenatal alcohol or cocaine exposure: The cumulative risk model predicted synergistic effects from exposure to multiple risk factors, and the fetal patterning of disease model predicted that prenatal insult will increase vulnerability to environmental risk. Four or more risks emerged as a threshold for poorer cognitive and behavioral outcome among the non-substance-exposed children, whereas substance-exposed children showed greater vulnerability to lower levels of environmental risk. Cumulative risk was associated with increased delinquent and internalizing behaviors only for the substance-exposed group. Results support the cumulative risk model for non-substance-exposed children and increased vulnerability to environmental risk among the substance-exposed group.

A wide range of factors in the social environment has been shown to put the child at risk for poorer cognitive and emotional development. One approach to understanding the effects of multiple risk exposure on children's outcomes has focused on the examination of cumulative risk. The cumulative risk model hypothesizes that the total number of risk factors to which the child is exposed is critically important in predicting child maladjustment (Garmezy, 1985; Rutter, 1979; Sameroff, Seifer, Baldwin, & Baldwin, 1993). This model suggests that, regardless of the specific risk factor, as the number of risk factors increases, child psychosocial adjustment becomes increasingly compromised. An alternative model for examining the impact of socioenvironmental risk focuses on the moderating impact of fetal patterning of adult disease. This approach is based on evidence from developmental biology and animal studies suggesting that environmental influences operating in utero can affect the propensity for a disease that may not become manifest until later in development (Gluckman & Hanson, 2004; Wadhwa, 2005). It is hypothesized that an intrauterine event may alter neural development so as to render certain individuals at risk for pathology later in life, increasing vulnerability to chronic illness, including depression, anxiety disorders, schizophrenia, and drug abuse (McDonald, Devan, & Hong, 2004). One implication of this model is that an intrauterine insult, such as fetal alcohol or drug exposure, might increase vulnerability to socioenvironmental risk factors encountered by a child who is reared in a less than optimal environment. The study reported here examined the impact of socioenvironmental risk on cognitive and behavioral outcomes from the perspectives of both models—cumulative risk and fetal patterning of disease—using data from a prospective longitudinal study of inner-city African American children.

Cumulative Risk

A cumulative risk model provides information regarding the number of risk factors the child can "tolerate" before his or her psychosocial adjustment deteriorates markedly (Jones, Forehand, Brody, & Armistead, 2002). For example, in the Isle of Wight study, when using a risk score based on the presence or absence of up to six risk factors (e.g., low socioeconomic status [SES], overcrowding, maternal

---

Chie Yumoto is now at the Department of Psychiatry, Harvard Medical School.

This research was funded by Grants R01-AA06966, R01-AA09524, and P50-AA07606 from the National Institute on Alcohol Abuse and Alcoholism, with supplemental support from a Minority Biomedical Research Support Grant S06-RR08167 from the National Institutes of Health, and a grant from the Joseph Young Sr. Fund from the State of Michigan. We thank Robert J. Sokol, Susan Martier, Joel Ager, and Erawati Bawle, our collaborators on this research; Douglas Barnett, for his suggestions; Neil Dodge, Lisa Chiodo, Renee Sun, Jeannine Tell, Brenda Tuttle, Jennifer Jester, and Raluca Corobana, who helped collect and analyze the data; and the mothers/caregivers and children who participated in the study. This study is based on a master's thesis completed by C.Y. at Wayne State University.

Correspondence concerning this article should be addressed to Sandra W. Jacobson, Department of Psychiatry and Behavioral Neurosciences, Wayne State University, 2751 East Jefferson, Room 460, Detroit, MI 48207. Electronic mail may be sent to sandra.jacobson@wayne.edu.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7906-0016

psychiatric disorder), the incidence of child psychiatric diagnosis increased sevenfold when the total risk score was 4 or more (Rutter, 1979). Werner and Smith (1982) used a set of 12 risk variables, including poverty, parental psychopathology, and disruption of family unit, and found that children with risk scores of 4 or more at 2 years of age had significantly poorer quality of adaptation at ages 10 and 18 than children with fewer than 4 factors. Applying this model to cognitive outcomes, Sameroff et al. (1993) used a cumulative risk index in their longitudinal study of IQ scores in 4- and 13-year-old children whose mothers were recruited to overrepresent maternal mental illness. The authors examined 10 risk factors (minority group status, occupation of head of household, maternal education, family size, father absence, stressful life events, parenting perspectives, maternal anxiety and mental health, and mother – child interaction—each of which was dichotomized as high or low risk), and a risk index score was calculated based on the number of risk factors in the child's family. They found that IQ scores decreased at both ages as the number of risks increased, with the risk index accounting for 34% of the variance in IQ at 4 years and 37% at 13 years. This effect persisted even after SES was controlled statistically and was still substantial after adjustment for the effects of maternal IQ.

Burchinal, Roberts, Hooper, and Zeisel (2000) compared the predictive power of the cumulative risk approach, in which each risk factor is dichotomized to permit tabulation of the number of high risks to which the child is exposed, with a multiple regression approach, in which child outcomes are predicted from individual risk variables examined simultaneously. The authors note that the multiple regression approach retains the full predictive power of the original continuous risk measures and permits evaluation of the relative importance of each predictor. An important advantage of the cumulative risk approach is that it permits examination of the hypothesis that the impact of multiple risks becomes exacerbated after the number of risks to which the child is exposed exceeds a given threshold.

African Americans have been underrepresented in studies examining cumulative risk, and some studies that have examined cumulative risk in African American children have yielded different findings from those studying Caucasian children. Deater-Deckard, Dodge, Bates, and Pettit (1998) found that individual differences in the increase in externalizing behaviors between 5 and 10 years of age were related to both individual risk factors and cumulative risk. However, their results suggested that most of the prediction of externalizing behavior from cumulative risk was found among Caucasian rather than African American children. Similarly, Gerard and Buehler (2004) found that the association of their cumulative risk measure with conduct problems in 7th- to 11th-grade children was stronger for Caucasian than for African American youth. In the Sameroff et al. (1993) study, the effect of the cumulative risk at age 4 years was also found only among the Caucasian families, although it was evident within both the Caucasian and the African American families at 13 years.

An improved understanding of vulnerability to socioenvironmental risk in the African American population is particularly important given that this population is disproportionately exposed to environments with high levels of psychosocial risk (McLoyd, 1990). In a study of 277 African American children living in single-mother households, Jones et al. (2002) found that cumulative risk was not related to externalizing problems but was predictive of internalizing problems. Among the individual risk factors that these authors examined, only inadequate parenting was significantly related to externalizing problems, whereas maternal depressive mood and inadequate parenting were related to internalizing problems. Kim and Brody (2005) examined risk factors longitudinally using the same data set. They found that an accumulation of family risk factors at 11 years of age was linked with maternal psychological functioning at 12 years of age, which forecast competence-promoting parenting practices at 13 years of age. These parenting practices indirectly predicted externalizing and internalizing behaviors 2 years later, a process that was mediated by youth self-regulation at 14 years of age. Thus, cumulative risk impacted the child's later behavior by influencing the child's psychosocial experiences, including interaction with his or her parents.

Fetal Origins of Vulnerability to Socioenvironmental Risk

Numerous studies have examined the relation of prenatal exposure to alcohol and illicit drugs to cognitive and emotional development. Fetal alcohol spectrum disorder (FASD) ranges from the most severe form—fetal alcohol syndrome (FAS), which is characterized by a distinctive craniofacial dysmorphology, small head circumference, and pre- and/or postnatal growth retardation (Hoyme et al., 2005)—to alcohol-related neurodevelopmental disorder (ARND), a term applied to children exposed prenatally to alcohol who lack the characteristic FAS dysmorphology but exhibit measurable, albeit generally subtler, neurobehavioral deficits (Hoyme et al., 2005; Stratton, Howe, & Battaglia, 1996). Although many children with FAS are mentally retarded (IQ < 70), most children with ARND perform in the normal range (Goldschmidt, Richardson, Stoffer, Geva, & Day, 1996; Mattson, Riley, Gramling, Delis, & Jones, 1997; Streissguth, Barr, & Sampson, 1990; Streissguth et al., 1991). These effects are generally more severe in children whose mothers were older at time of delivery and/or had a history of alcohol abuse problems and less severe in those who received more optimal parental intellectual stimulation and emotional support (S. W. Jacobson, Jacobson, Sokol, Chiodo, & Corobana, 2004). In addition to its effects on cognitive development, prenatal alcohol exposure has been linked to depression in children (O'Connor et al., 2002) and adults (Carmichael, Feldman, Streissguth, Sampson, & Bookstein, 1998; Famy, Streissguth, & Unis, 1998; Spohr & Willms, 1993; Steinhausen, Willms, & Spohr, 1993).

Because the biological insult associated with prenatal substance exposure is likely to be confounded with environmental influences associated with maternal drinking during childhood, it is important to try to distinguish these influences. O'Connor and Paley (2006) found that whereas prenatal alcohol exposure was associated with increased depressive symptoms in the child, these symptoms were not related to current maternal drinking patterns or depression, supporting the hypothesis that fetal exposure plays a critical role in increasing vulnerability to child depression. Similarly, S. W. Jacobson et al. (2006) found that prenatal alcohol exposure was predictive of teacher reports of externalizing problems in the classroom but that these problems were unrelated to current maternal drinking, again suggesting a particular role for fetal exposure in increasing the risk of developing emotional problems in childhood.

Prenatal exposure to cocaine has also been studied in relation to childhood cognitive and emotional development. Coles and Black (2006) point out that, by contrast to fetal alcohol exposure, findings on the effects of prenatal cocaine exposure have been mixed, with the clearest effects seen in cases of heavy cocaine exposure (Alessandri & Bendersky, 1998; Frank & McCarten, 1999; S. W. Jacobson, Jacobson, Sokol, Martier, & Chiodo, 1996; Singer et al., 2001; Singer et al., 2002). There have been relatively few reports of adverse effects on cognitive performance, play behavior, academic achievement, and classroom behavior (Frank et al. 2002; Mayes, 2002; Mayes, Molfese, Key, & Hunter, 2005). The effects have been found primarily on behavioral problems, task persistence and attention, and language performance. Bendersky, Bennett, and Lewis (2006) examined childhood aggression at 5 years of age in a multiple risk model that included cocaine exposure, environmental risk, and gender as predictors and found that all three of these factors were related to a composite aggression score based on reports from multiple informants. Carta et al. (2001) examined the effects of cumulative prenatal substance exposure and cumulative environmental risk on the developmental trajectories of 278 infants, toddlers, and preschool children. Both prenatal exposure and environmental risk added unique variance to the prediction of rate of growth and developmental level and after controlling for birth weight and gender.

The studies to date have focused primarily on detecting main effects of prenatal substance exposure on long-term outcome after controlling for the influences of the social environment. However, studies on preterm infants have demonstrated that fetal and neonatal insult can also increase vulnerability to environmental risk. For example, Levy-Shiff, Einat, Mogilner, Lerman, and Krikler (1994) examined the role of biological and psychosocial factors in predicting long-term outcome in adolescents who were born prematurely at very low birth weight compared with those who were born full term at normal birth weight. They found that SES predicted visual–motor coordination and hyperactive behavior in the very low-birth-weight group but not in the normal-birth-weight group, indicating an interaction between biological vulnerability and environmental risk. Landry, Smith, Swank, Assel, and Vellet (2001) studied the role of early versus ongoing maternal responsiveness in predicting cognitive development in full-term and preterm children at five ages ranging from 6 months to 4 years. They found that both preterm and full-term children's developmental rates in cognitive skills were close to the expected gains of 12 months per year with consistently high levels of responsive parenting. In contrast, preterm children's gains were only 9 months per year, compared to 11 months per year gained by full-term children, when parented by consistently low-responsive mothers. Thus, their data demonstrate the greater vulnerability of children who have experienced an early biological insult to less than optimal parental responsiveness (see also Parker, Greer, & Zuckerman, 1988). Whereas most of these studies have focused on specific environmental risks, the few that have examined the impact of neonatal insult on increased vulnerability to cumulative risk lacked an appropriate control group (e.g., Escalona, 1982). Based on these studies from the preterm literature, we hypothesize that children who are biologically vulnerable as a result of prenatal alcohol or drug exposure and are also exposed to cumulative environmental risks may also be at "double jeopardy" for poor outcomes (Carta et al., 2001).

Aims of the Study

The principal aim of this study was to examine the degree to which cumulative risk and fetal substance exposure modify vulnerability to socioenvironmental risk in a sample of economically disadvantaged, African American inner-city children. We looked at socioenvironmental risks associated with poor outcome and compared their cumulative effects on the child's cognitive and emotional development in prenatally exposed and nonexposed groups. The goal was to examine whether cumulative risk operates differently in these two groups, as well as the degree to which the impact of cumulative risk in the nonexposed African American children resembles that reported in studies of Caucasian children. In addition, it was hypothesized that children who were prenatally exposed would be more vulnerable to cumulative risk factors. Specifically, we predicted that the magnitude of association between risk factors and negative outcome would be stronger in exposed children and that the threshold for a poorer outcome would be lower among children who were exposed to substances prenatally than those who were not.

Method

Participants

Mothers were recruited during pregnancy for a longitudinal study on the effects of prenatal exposure to alcohol and cocaine on child development (S. W. Jacobson et al., 2004; S. W. Jacobson et al., 2006). All African American women were screened for alcohol consumption during their first visit to a prenatal clinic in a large urban maternity hospital ($M = 23.4$ weeks of gestation, $SD = 7.9$) from 1986 to 1989 using a timeline follow-back interview (S. W. Jacobson, Chiodo, Jacobson, & Sokol, 2002). Those averaging seven drinks per week or more at the time of conception (≥ 0.5 oz absolute alcohol [AA] per day), the lowest level at which adverse effects have been observed (J. L. Jacobson & Jacobson, 1994; J. L. Jacobson, Jacobson, Sokol, & Ager, 1998), were invited to participate in the study in order to overrepresent women who were drinking at moderate to heavy levels during pregnancy. A random sample of 5% of the abstainers and lighter drinkers (< 0.5 oz AA/day) from the clinic population were also invited to participate. In addition, the sample includes 78 women who reported heavy cocaine (≥ 2 days per week), light alcohol ($<$ seven drinks per week) use at the time of recruitment. The inclusion of this subgroup substantially reduced the confounding of prenatal alcohol and cocaine exposure (S. W. Jacobson et al., 1996); only 8 moderate to heavy drinkers reported heavy use of cocaine at recruitment. Mean maternal age at recruitment was 26.4 years ($SD = 6.0$), with 6.3% between 14 and 18 years. Infant exclusionary criteria included birth weight $< 1{,}500$ g, gestational age < 32 weeks, major chromosomal anomalies or neural tube defects, or multiple births. A total of 480 children and their mothers were recruited and assessed in the laboratory during infancy.

The sample evaluated in this study comprised the 337 children and their mothers or current primary caregivers who were assessed at 7.5 years of age ($M = 7.8$, $SD = 0.3$). A comparison of the 7.5-year study participants (70.0%) with those lost to follow-up (30.0%) indicated no differences in SES, maternal educational level, alcohol consumption at conception, drug use during pregnancy, or psychopathology (all $ps > .20$). Retention rates were similar in the exposed (72.1%) and nonexposed (66.3%) groups, $\chi^2(1) = 1.60$, *ns*. The only significant difference was that the mothers who participated in the follow-up were slightly older ($M = 26.7$) compared with 25.5 years for nonparticipants, $t(478) = -2.07$, $p < .05$.

To examine the degree to which the impact of socioenvironmental risk may differ depending on prenatal exposure history, two groups were created based on prenatal alcohol and cocaine exposure. The "exposed" group consisted of 227 children whose mothers reported consuming ≥ 0.5 oz AA/day (the equivalent of one or more standard drinks/day) at time of conception or using cocaine at least once per week during pregnancy. The "nonexposed" group consisted of 110 children whose mothers drank < 0.5 oz AA/day with no or, at most, rare, minimal cocaine use during pregnancy. Most of the prenatal alcohol exposure in the exposed group was in the moderate range ($M = 1.3$ oz AA/day, $SD = 2.3$), and only 3 children met criteria for a diagnosis of FAS. When averaged across pregnancy, cocaine was used regularly (≥ 1 day/week) by 7.4% of the mothers.

At the 7.5-year visit, the biological mother was the primary caregiver for a large majority of the children (see Table 1). The primary caregivers for the remaining children were the father (1.3%) or a family relative, usually a grandmother (8.5%) or aunt (1.3%); 1 child (0.3%) had been adopted by foster parents. Median SES for both the exposed and nonexposed groups fell in the fourth level on the Hollingshead (1975) scale, the second lowest of five levels on the scale.

Table 1
Sample Characteristics

	Nonexposed (n = 110)		Exposed (n = 227)		
	M or %	*SD*	*M* or %	*SD*	*t* or χ^2
Maternal/primary caregiver					
Socioeconomic status[a,b]	26.5	9.7	25.4	11.7	0.9
Group 1 (%)	0.0		1.3		
Group 2 (%)	13.6		15.0		
Group 3 (%)	21.8		14.5		
Group 4 (%)	38.2		30.0		
Group 5 (%)	26.4		39.2		
Age[b]	33.3	9.4	37.8	9.6	−4.1**
Parity[c]	1.0	1.2	1.5	1.6	−3.7**
Verbal competence[c,d]	74.1	12.7	73.4	14.0	0.4
Marital status (% married)[b]	25.5		15.9		3.8*
Welfare (% receiving)[b]	63.6		72.2		2.6
Child					
Primary caregiver (% biological mother)	90.0		83.3		2.7
Gender (% male)	49.1		62.6		5.5*

[a]Hollingshead Four Factor Index of Social Status Scale (1975). [b]Primary caregiver. [c]Biological mother. [d]Peabody Picture Vocabulary Test–Revised (Dunn & Dunn, 1981).
$*p < .05$. $**p < .01$.

Differences between the nonexposed and exposed groups were observed on four demographic measures. The mothers of the exposed children were older, higher in parity, and less likely to be married, and the exposed children were more likely to be male. Because the difference in parity was a consequence of the difference in primary caregiver age, parity was not considered further. Caregiver age and marital status and child gender were included in the regression analyses described in the following, where appropriate, to control for their potential confounding effects.

Procedures

All child assessments were conducted by examiners who were blind with respect to the child's prenatal exposure and risk status. Procedures were approved by the Wayne State University Human Investigation Committee. Written informed consent was obtained from the mother/primary caregiver at recruitment and at the 7.5-year assessment, and oral assent from the child at 7.5 years. Child cognitive performance was assessed in the laboratory, and behavior problems were rated by the child's classroom teacher. The primary caregiver was interviewed in detail about her current alcohol and drug consumption and other socioenvironmental factors and was administered a test of verbal competence and a test assessing severity of alcohol-related problems. The primary caregiver received monetary compensation and a photo of the child, and the children received a small gift for their participation.

Assessment Battery

Eight risk variables in the child's social and family environment were included in the analyses reported here. Six of these constructs were adapted from Sameroff et al.'s (1993) study, which is a landmark study in examining the effects of cumulative risk on child development. Given the focus of this study on prenatal substance abuse exposure, current substance abuse and parental violence, which often accompanies substance abuse, were also included in the analyses. In order to calculate the cumulative risk index, each of these variables was dichotomized using the criteria listed in the following. Missing data were estimated by using the group median of the exposure group to which the participant belonged. Table 2 summarizes the risk cutoff criteria and presents the percentage of each group that was categorized as high risk.

SES. Numerous studies have demonstrated that lower SES is associated with a wide range of poorer outcomes in children, including cognitive and socioemotional functioning (e.g., Bradley & Corwyn, 2002) and neurocognitive abilities (e.g., Noble, McCandliss, & Farah, 2007). In this study, SES was computed on the Hollingshead (1975) scale separately for the

Table 2

Risk Factors and Cutoff Scores

			% in risk group			Coefficient of variation	
	Measure	Risk cutoff	Nonexposed	Exposed	χ^2	Nonexposed	Exposed
Socioeconomic status	Hollingshead scale[a]	< 20	25.5	39.2	6.18*	0.36	0.46
Number of children	—	≥ 4	13.6	11.9	0.21	0.93	0.89
Presence of father	Lives with biological father	Father absent	82.7	80.2	0.31		
Stressful life events	Life Events Scale[b]	Highest quartile	22.7	28.2	1.14	0.78	0.77
Depression	Beck Depression Inventory[c]	≥ 16	11.8	18.5	2.42	1.00	0.83
Parent–child interaction	HOME[d]	Lowest quartile	26.4	26.0	0.01		
Parental violence	Conflict Tactics Scale[e]	95 percentile U.S. norm	39.1	43.6	0.62	1.88	1.63
Current substance abuse	MAST[f]	≥ 5	20.9	50.7	27.12**		
	or cocaine use	≥ 1 day/week					
	or alcohol use	≥ 4 drinks/day					

[a] Hollingshead Four Factor Index of Social Status Scale (1975). [b] Adapted from Holmes and Rahe (1967). [c] Beck and Steer (1987). [d] Home Observation for Measurement of the Environment (Caldwell & Bradley, 1984). [e] Straus (1979). [f] Michigan Alcoholism Screening Test (Selzer, 1971).

*$p < .05$. **$p < .01$.

primary caregiver and partner based on their educational attainment and occupational status, which was averaged together if both were employed; if not, SES was based on the educational attainment and occupational status of the head of household. The continuous Hollingshead SES measure ranges from 8 to 66, which is then grouped into five SES levels; when SES was dichotomized for risk categorization, a score < 20—the lowest of the five Hollingshead SES levels—was classified as high risk.

Number of children at home. It has been shown that having a larger number of children in the home is associated with poorer child cognitive outcome (Luster & McAdoo, 1994). In our study, data on number of children younger than 18 years living in the household were obtained. Following Sameroff et al. (1993), four or more children was considered as high risk.

Presence of father at home. The primary caregiver was asked whether the child lives with his or her father. Although we obtained information regarding the presence of biological fathers, stepfathers, and any father figures (e.g., grandfathers, uncles), the absence of the biological father was used as the high-risk category criterion in this study based on our finding that only the presence of the biological father protected against psychosocial risk in this population (Yumoto et al., 2002). In this sample, it was common for children to have second households where they spend weekends or holidays. However, the biological father was considered to be "present" only when the child lived with him in the primary household.

Stressful life events. Negative life events experienced by primary caregivers (e.g., job loss, divorce, or illness) have been found to be related to children's depression (Compas, Howell, Phares, Williams, & Giunta, 1989), delinquency (Tolan, 1988), aggression (Attar, Guerra, & Tolan, 1994), and cognitive performance (Krishnakumar & Black, 2002). Information was obtained using an adaptive version of the Schedule of Recent Events (Holmes & Rahe, 1967), on which primary caregivers rated any of 43 events they experienced during the preceding year in terms of how stressful they found each event. The continuous measure of life stress consisted of the sum of these ratings; a score in the highest quartile was used as the risk cutoff. One case was missing data on the Schedule of Recent Events and was estimated by using a group median.

Depression. Maternal depression has been shown to be related to children's behavior problems, including aggression, anxiety, and social incompetence (Dodge, 1990) and externalizing problem behaviors (Krishnakumar & Black, 2002). To assess maternal/primary caregiver depression, the present study used the Beck Depression Inventory (BDI; Beck & Steer, 1987), which was also highly correlated with the Global Severity Index on the Symptoms Checklist 90–Revised (Derogatis, 1992) in this sample ($r = .73$, $p < .001$). We used the BDI score of 16 or above, which is considered indicative of moderate to severe depression, as the high-risk cutoff.

Parent–child interaction. Quality of the home environment, as measured by the elementary age version of the Home Observation for Measurement of the Environment (HOME; Caldwell & Bradley, 1984), has been found to be related to child cognitive functioning among 6- to 9-year-old African American children (Luster & McAdoo, 1994). In addition, emotional quality of the parent–child relationship on the elementary age HOME was significantly related to both internalizing and externalizing behaviors in children aged 6–10 years (Wasserman, Miller, Pinner, & Jaramillo, 1996). In our study, scores on two subscales from the HOME—verbal responsivity and emotional climate—were used to assess quality of primary caregiver–child interaction. The parent–child interaction variable was constructed by combining the standardized scores on these two subscales. Both subscales comprised largely examiner ratings based on the observation of the emotional quality of interaction between the primary caregiver and the child. Although mean interobserver reliability was not available for these specific items at this age, previous interobserver reliability on the HOME for this cohort averaged 95% (range = 88%–100%; S. W. Jacobson, Jacobson, Sokol, Martier, & Ager, 1993) and the HOME data obtained at this age were predictive of child IQ (S. W. Jacobson et al., 2004), as reported in other studies. A score in the lowest quartile of the sample distribution, indicating low sensitivity and cold emotional climate at home, was used to define the risk category. For two missing cases, the group median was used to estimate missing values.

Parental violence. Exposure to domestic violence, as measured by the Conflict Tactics Scale (CTS; Straus, 1979), has been linked to increased internalizing and aggressive behavior problems (Holden & Ritchie, 1991) and increased general psychopathology among 6- to 12-year-olds (McCloskey, Figueredo, & Koss, 1995). The CTS assesses the incidence and severity of verbal and physical conflict between the primary caregiver and his or her partner. Scores on the CTS violence scale falling in the top 5% of the national norms for this scale were considered high risk.

Primary caregiver's current substance abuse. Parental substance abuse has been considered a significant risk factor that negatively impacts on children's cognitive functioning and is related to increased externalizing and internalizing behaviors (Moss, Vanyukov, Majumder, Kirisci, & Tarter, 1995; Nunes et al., 1998; Puttler, Zucker, Fitzgerald, & Bingham, 1998; Weinberg, 1997; Wilens et al., 2002). We assessed current substance abuse by the primary caregiver based on the following three measures: (a) frequency of alcohol use, (b) frequency of opiate and cocaine use, and (c) Michigan Alcoholism Screening Test (MAST; Selzer, 1971) score. The MAST is a 25-item self-report measure widely used to identify problems related to alcohol consumption, such as alcohol-related loss of job or friendship or arrest for driving while intoxicated. A positive MAST score (≥ 5) indicates that the respondent has experienced psychosocial and/or physiological problems related to alcohol abuse and is alcoholic or at risk for alcoholism. The continuous variable used to measure current substance abuse was created by averaging standardized scores for ounces of AA per day, days per month of opiate and cocaine use, and the MAST score obtained at the 7.5-year follow-up assessment. For six cases that were missing one or two of the aforementioned variables, the median of the standardized scores for each exposure group was used to estimate missing values. The high-risk group for current substance abuse comprised primary caregivers who currently drank at least 2 oz AA (four or more standard drinks)/day, currently used cocaine once/week or more, or scored 5 or more on the MAST.

Table 2 presents the coefficients of variation (CV; *SD/M*) for the eight risk factors in the nonexposed and exposed groups to assess potential restriction of range. Of the five nonstandardized continuous socioenvironmental risk variables, CV for the exposed group was slightly larger on SES (0.5, compared with 0.4), slightly smaller on maternal depression (0.8, compared with 1.0) and parental violence (1.6, compared with 1.9), and very similar on the remaining measures. With regard to the outcome measures, CV for the exposed group was very similar for the three IQ measures, slightly lower for externalizing behavior (1.1, compared with 1.2) and slightly higher for internalizing behavior (1.1, compared with 1.0).

In addition to the eight socioenvironmental risk factors, verbal competence of the child's biological mother was assessed on the Peabody Picture Vocabulary Test–Revised (PPVT–R; Dunn & Dunn, 1981), which is strongly correlated with Wechsler Adult Intelligence Scale full-scale IQ ($r = .72$) and has been validated for use with this cohort (S. W. Jacobson, Jacobson, & Frye, 1991). PPVT–R score was used as a covariate to test the degree to which the socioenvironmental risk factors affect cognitive development over and above the contribution of maternal verbal intellectual competence.

Outcome Variables

The Wechsler Intelligence Scale for Children–Third Edition (WISC–III; Wechsler, 1991) was used to assess the child's cognitive functioning. The WISC–III yields full-scale IQ (FSIQ), verbal IQ (VIQ), and performance IQ (PIQ) scores. The child's classroom teacher was sent the Teacher Report Form (TRF; Achenbach, 1991) to complete and received a small gift when the questionnaire was returned. Four summary measures derived from the TRF were used in the analyses: aggressive behavior problems; delinquent behavior problems; externalizing total score, which comprises the aggressive and delinquent behavior problem scales; and internalizing total score, which consists of the withdrawn, somatic complaints, and anxious depressed scales.

Data Analyses

Dichotomous risk variables were constructed from the social and family variables using the aforementioned criteria. A cumulative risk index was then computed by assigning a score of 1 for the presence of each risk factor. Cumulative risk was defined as the total number of risk factors of the eight examined in this study.

All analyses were conducted separately for the substance-exposed and nonexposed groups. Multiple regression analyses were conducted using the eight risk variables as continuous measures to examine the associations of social and family variables with the child outcome variables. Hierarchical regression analyses were then conducted to control for three additional sociodemographic potential confounding variables—primary caregiver's age and marital status and child gender—by including all the risk variables in the first step and the additional potential confounders in the second step.

Following these analyses, a correlational analysis was conducted to examine the relation of cumulative risk to the outcome variables. VIQ and PIQ scores were then examined in relation to the cumulative risk score in analyses of covariance (ANCOVAs) to detect whether there were thresholds above which performance is affected particularly strongly, after controlling for the biological mother's verbal competence. The cumulative risk index was recoded into four groups (0–1, 2, 3, and 4 or more) to ensure a sufficient number of cases in each group. Analyses of variance were conducted to assess whether there were threshold effects in the relation of cumulative risk to the behavior problem outcomes assessed. Post hoc comparisons were based on least square difference (LSD) analyses.

Results

Table 3 presents correlations among eight risk factors for each group. The highest correlation among the risk factors in both the nonexposed and the exposed

Table 3
Correlations Among Risk Factors

	Socioeconomic status (SES)	Number of children	Presence of father	Stressful life events	Depression	Parent–child interaction	Parental violence	Current substance use
SES	—	−.23*	.12	.14	−.02	.10	.02	.02
Number of children	−.05	—	.18†	.10	.05	−.18†	−.05	−.02
Presence of father	.17*	.06	—	.09	−.08	.10	−.02	.01
Stressful life events	.02	.14*	.09	—	.48***	−.08	.30**	.21*
Depression	−.09	.06	.06	.54***	—	−.06	.29**	.11
Parent–child interaction	.25***	−.07	.03	−.05	−.20**	—	.02	−.04
Parental violence	−.12†	.11†	.09	.31***	.31***	−.11†	—	.30**
Current substance abuse	−.08	−.03	.13†	.33***	.32***	−.22**	.31***	—

Note. Values are Pearson's *r*s. Values below the diagonal are from the exposed group and above the diagonal are from the nonexposed group.
$^\dagger p < .10$. $^*p < .05$. $^{**}p < .01$. $^{***}p < .001$.

groups was between the primary caregiver's depression and stressful life events in the past year. In the nonexposed group, the next highest correlations were between stressful life events and parental violence and between current substance abuse problems and parental violence. In the exposed group, the next highest correlation was between current substance abuse problems and stressful life events. In the exposed group, SES, primary caregiver's depression, and primary caregiver's current substance abuse problem were all significantly correlated with the responsiveness and emotional climate at home, whereas none of these risk factors were associated with quality of home environment in the nonexposed group. Another difference observed between two groups was in the relation between the primary caregiver's depression and substance abuse, which was significant only in the exposed group.

When the risk factors were examined as continuous measures, the nonexposed and exposed groups differed on only three. Mothers of the substance-exposed children were more likely to have current substance abuse problems, $t(335) = -7.72$, $p < .001$; be depressed, $t(335) = -2.68$, $p < .01$; and report more stressful life events in the previous year, $t(335) = -2.14$, $p < .05$. Although we have previously reported effects of prenatal alcohol exposure on IQ (S. W. Jacobson et al., 2004) and TRF behavior problems (S. W. Jacobson et al., 2006), in these analyses, no differences in mean IQ, delinquency, aggression, or internalizing or externalizing behavior problem scores were found between the nonexposed and exposed groups because effects of prenatal alcohol and cocaine exposure were combined in this study of environmental risk, thereby presumably masking their specific teratological impact (see Table 4).

Table 4
Mean IQ and TRF Scores (SDs) by Exposure Group

	Nonexposed	Exposed	*t* value
WISC-III[a]			
Full-scale IQ	85.6 (13.1)	83.6 (11.9)	1.4
Verbal IQ	89.0 (12.9)	86.6 (12.1)	1.6
Performance IQ	84.5 (13.6)	83.2 (12.7)	0.9
TRF[b]			
Delinquency	1.9 (2.4)	2.4 (2.8)	−1.6
Aggression	8.0 (10.2)	10.3 (11.8)	−1.5
Externalizing	9.9 (12.0)	12.7 (14.1)	−1.6
Internalizing	7.1 (6.8)	6.1 (6.4)	1.3

Note. WISC−III = Wechsler Intelligence Scale for Children−Third Edition; TRF = Teacher Report Form.
[a]*ns* = 110 for nonexposed and 227 for exposed groups. [b]*ns* = 80 for nonexposed and 171 for exposed groups.

Multiple Regression Analyses

Table 5 summarizes the results from the multiple regression analyses. The simple correlations relating the cumulative risk index score with the outcome variables are shown in the bottom row. The principal predictors of FSIQ for the nonexposed group were number of children living at home and parental violence, with an effect of SES falling short of significance. For the children in the exposed group, SES and parental current substance abuse were the principal predictors of FSIQ. Similar patterns were observed with VIQ and PIQ, except that having a biological father living at home was negatively related to PIQ in the exposed group, $\beta = -.14$, $p = .04$, and, for the nonexposed group, parental violence was related to VIQ, $\beta = -.25$, $p = .01$, but not to PIQ. The multiple *R* based on the eight discrete risk factors measured as continuous variables was markedly stronger than the *r* for cumulative risk index for both groups. Moreover, in each regression, the multiple *R* was markedly stronger than the individual standardized regression coefficients, indicating that the combined effect of those risk factors measured as continuous variables exceeds that associated with any of the individual risk factors.

Differential patterns of risk for the TRF scores were also observed for the nonexposed and exposed groups, with one notable exception. The primary caregiver's responsivity and emotional climate at home predicted externalizing behavior problems (i.e., delinquent and aggressive behaviors, and the externalizing total score) for both groups. Whereas the primary caregiver's depression was related to the externalizing behaviors in the nonexposed group, the primary caregiver's substance use problems were related to delinquent behavior problems only for the children in the exposed group. Internalizing behavior problems were not related to any of the individual risk factors, and the multiple *R* for the regression relating the eight risk factors to internalizing total score was not significant for either group.

The regression analyses shown in Table 5 were repeated to control for primary caregiver's age, marital status, and the gender of the child, to examine whether those demographic variables alter the effects observed in the regression. The results were essentially unchanged, except that the effects of the primary caregiver's responsivity and emotional climate at home on delinquent and aggressive behaviors were reduced somewhat in the nonexposed group when the gender of the child was included in the analyses (externalizing total score was reduced from $\beta = -.26$ to $-.18$; delinquent behavior score, from $\beta = -.24$ to

Table 5
Full-Scale IQ and Teacher Report Form Scores in Nonexposed Versus Exposed Groups

	Full-scale IQ				Externalizing total score				Delinquent behavior				Aggressive behavior			
	Nonexposed (n = 110)		Exposed (n = 227)		Nonexposed (n = 80)		Exposed (n = 171)		Nonexposed (n = 80)		Exposed (n = 171)		Nonexposed (n = 80)		Exposed (n = 171)	
Groups	r	B	r	B	r	β	r	β	r	β	r	β	r	β	r	β
Socioeconomic status	.25**	.18†	.23**	.22**	.05	.09	.02	.06	−.06	.01	−.09	−.03	.07	.10	.05	.08
Number of children	−.28**	−.24**	−.10†	−.10	.03	.03	.05	.05	.13	.14	.10†	.09	.00	−.00	.03	.03
Presence of father	−.11	−.11	−.10†	−.12†	−.12	−.16	−.11†	−.12	−.15†	−.20†	−.12†	−.13†	−.11	−.15	−.11†	−.12
Stressful life events	.11	.16	.02	.04	−.03	.06	.00	.04	−.09	−.03	.06	.03	−.01	.07	−.01	.04
Depression	.06	.06	.02	.10	−.18†	−.27*	−.06	−.10	−.16†	−.23†	.02	−.08	−.17†	−.26*	−.08	−.10
Parent–child interaction	.11	.08	.19**	.12†	−.22*	−.26*	−.15*	−.16*	−.24*	−.24*	−.25**	−.20*	−.20*	−.24*	−.12†	−.15†
Parental violence	−.13†	−.21*	−.05	.02	.03	.10	−.08	−.11	.01	.10	.04	−.07	.04	.10	−.11†	−.12
Current substance abuse	−.04	−.02	−.16**	−.15*	−.04	−.05	.07	.10	−.09	−.09	.24**	.23*	−.02	−.04	.03	.06
Multiple R	.42**		.35**		.36**		.25*		.39**		.36**		.34**		.24*	
Cumulative risk index	−.19**		−.15*		.06		.06		.12		.20**		.04		.03	

$^\dagger p < .10$. $^* p < .05$. $^{**} p < .01$.

−.20; aggressive behavior score, from β = −.24 to −.17), suggesting that the effect of quality of parenting on externalizing was possibly due, in part, to the fact that girls, who are less prone to exhibit externalizing behavior, were provided with more optimal interaction and warmer emotional climate at home in the nonexposed group.

Cumulative Risk Index

The cumulative risk index was calculated by counting number of risk factors the child experienced of the eight risk factors examined. Children in the exposed group experienced somewhat more risk factors (M = 3.0, range = 0–7) than those in the nonexposed group (M = 2.4, range = 0–6; t = 14.6, $p < .05$). The correlations of the cumulative risk index with FSIQ and the behavioral outcomes are shown in the bottom row of Table 5. As expected, a higher cumulative risk index was associated with lower FSIQ in both groups. Although the multiple R from the eight-variable regression was significantly related to all child outcome measures except the internalizing total score, the cumulative risk index was not related to the externalizing total score or the aggressive behavior score in either group. The cumulative risk index was related to the delinquent behavior score only in the exposed group. The internalizing total score was also related to the cumulative risk index score only in the exposed group (r = .20, $p < .01$, compared with r = −.02, *ns*, in the nonexposed group).

Threshold Effects

The ANCOVAs revealed different patterns for the exposed and nonexposed groups on VIQ (see Figure 1). The VIQ scores among the children who were not exposed to alcohol or drugs during pregnancy dropped dramatically in the presence of four or more risk socioenvironmental risk factors (LSD p = .05, for three vs. four or more risk factors), whereas the VIQ in the exposed group decreased gradually as the number of risk factors they had experienced increased (r = −.13, p = .06). By contrast, with the PIQ, both groups showed a gradual linear decrease as the number of risk factors increased (see Figure 2; rs = −.15, p = .10, and r = −.14, p = .04, for the nonexposed and exposed groups, respectively).

Differential patterns were also observed for the TRF scores. Figure 3 shows that delinquent behavior observed by the child's teacher significantly increased with four or more risk factors in the nonexposed group (LSD $p < .01$, for three vs. four or more risk

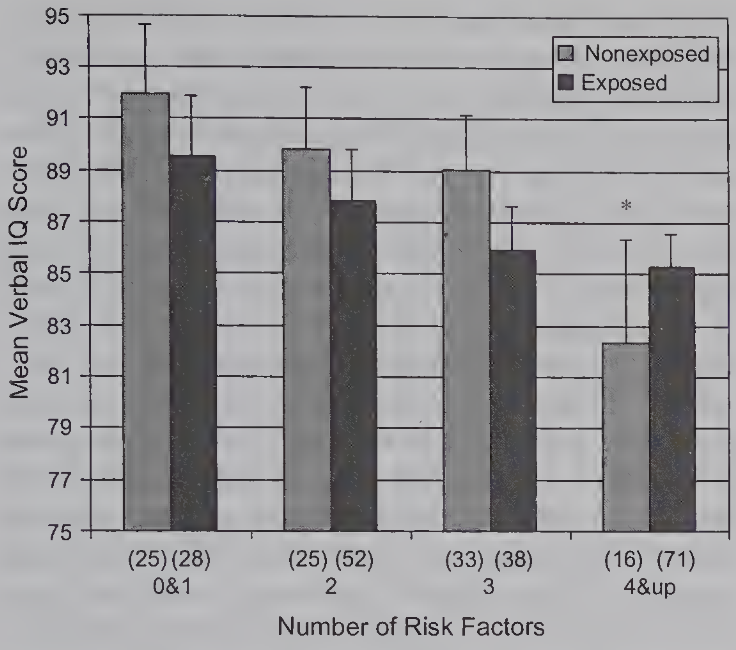


Figure 1. Mean verbal IQ score as a function of a number of risk factors in nonexposed and exposed groups.
Note. Numbers in parentheses are sample sizes for each subgroup. Vertical lines depict standard errors of the means.
*$p < .05$ based on least square difference post hoc test for three versus four or more risk factors in nonexposed group.

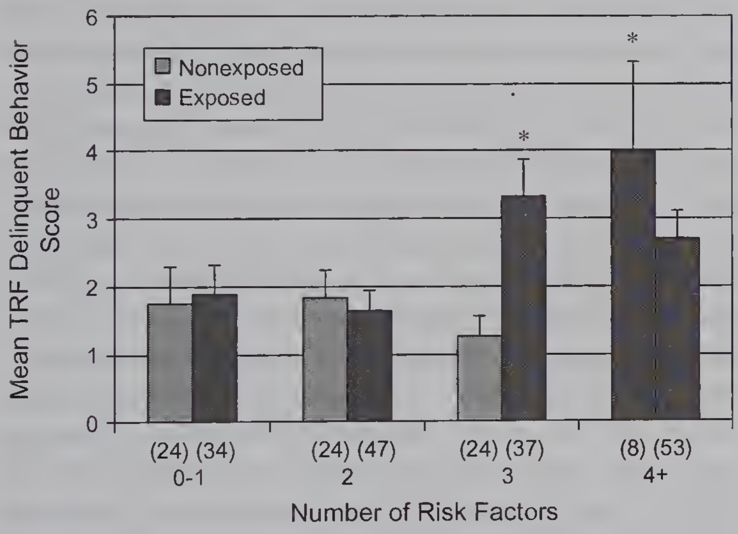


Figure 3. Mean Teacher Report Form (TRF) delinquent behavior score as a function of a number of risk factors in nonexposed and exposed groups.
Note. Numbers in parentheses are sample sizes for each subgroup. Vertical lines depict standard errors of the means.
*$p < .01$ based on least square difference post hoc test for three versus four or more risk factors in nonexposed group and for two versus three risk factors in exposed group.

factors). By contrast, delinquent behavior significantly increased after three risk factors in the exposed group (LSD $p < .01$, for two vs. three risk factors). No threshold effects were observed for aggressive behavior assessed by teachers in either group. As shown in Figure 4, the number of internalizing problems reported by teachers increased significantly with four or more risk factors in the nonexposed group (LSD $p = .03$, for three vs. four or more risk factors). In the exposed group, increasing risk was associated with a gradual increase in internalizing total score ($r = .20, p < .01$).

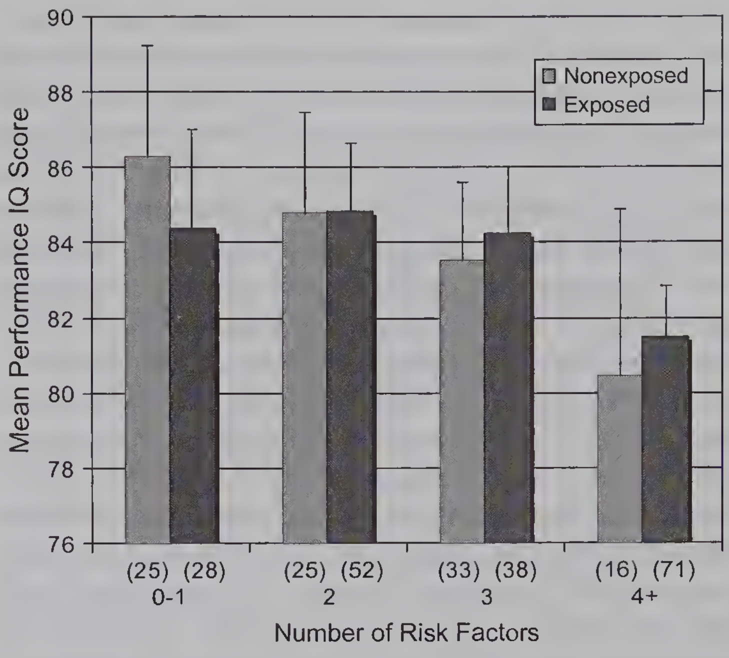


Figure 2. Mean performance IQ score as a function of a number of risk factors in nonexposed and exposed groups.
Note. Numbers in parentheses are sample sizes for each subgroup. Vertical lines depict standard errors of the means.

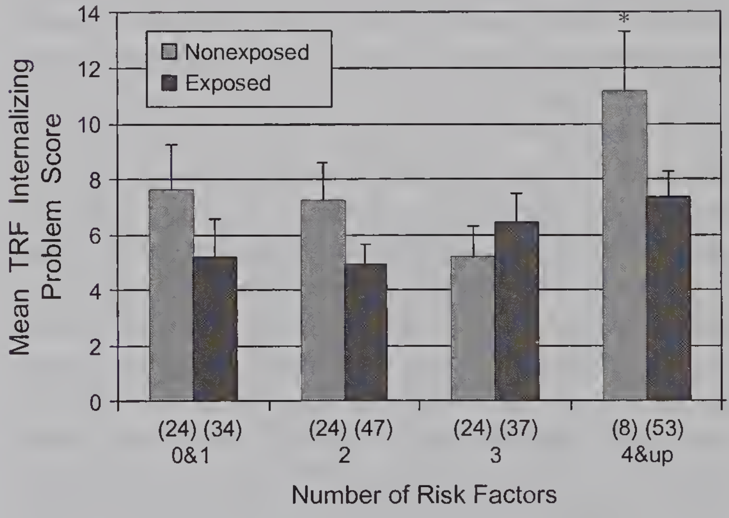


Figure 4. Mean Teacher Report Form (TRF) internalizing problem score as a function of a number of risk factors in nonexposed and exposed groups.
Note. Numbers in parentheses are sample sizes for each subgroup. Vertical lines depict standard errors of the means.
*$p < .05$ based on least square difference post hoc test for three versus four or more risk factors in nonexposed group.

Discussion

A major aim of this study was to examine the impact of prenatal exposure to alcohol and/or cocaine on vulnerability to risks encountered in the childrearing environment. When we previously examined the data looking specifically at moderate to heavy exposure to alcohol during pregnancy on IQ and externalizing behaviors, we reported significant effects of fetal alcohol exposure on IQ, working memory, number processing, and a range of behavioral endpoints on the TRF, using continuous measures of alcohol exposure (Burden, Jacobson, & Jacobson, 2005; Burden, Jacobson, Sokol, & Jacobson, 2005; S. W. Jacobson et al., 2003; S. W. Jacobson et al., 2004; S. W. Jacobson et al., 2006). These and other infant effects were *specific* to prenatal alcohol exposure (S. W. Jacobson, 1998) and different from those found for heavy cocaine or other teratogenic exposures (J. L. Jacobson et al., 1998) for this cohort. The lack of group differences reported here on these outcomes reflects the inclusion of cocaine-exposed children with little or no alcohol exposure in the high-exposure group in this study that was designed to focus on the moderating effects of prenatal substance exposure, thereby obscuring the prenatal alcohol effects. Thus, as expected, the prenatally exposed children as a group did not differ from the nonexposed group on the developmental outcomes examined—IQ and internalizing and externalizing behavior problems. Despite the similar mean IQ and behavior problem scores, however, the data revealed that differences in vulnerability to environmental risk operate in these two groups, thereby supporting the fetal patterning of later development hypothesis. Not surprisingly, current maternal substance abuse impacted more strongly on the IQ scores of the prenatally exposed children, presumably because of its more limited range in the nonexposed group. By contrast, number of children in the household was related to FSIQ only in the nonexposed group. Being raised in a smaller family had a beneficial effect on cognitive performance only for the children not exposed prenatally to alcohol or cocaine.

The findings from previous studies regarding the applicability of cumulative risk models of the impact of socioenvironmental factors on development in African American children have been inconsistent due, in part, to limited sample sizes. Our multiple regression data (Table 5) make clear, however, that the joint effect of multiple environmental risks markedly exceeds that of any of the individual risk factors. In contrast, the cumulative risk index, which is based on dichotomizing each of the risk factors and is, therefore, generally less sensitive (Burchinal et al., 2000), was significant for the cognitive outcomes for both the substance-exposed and the nonexposed children. For the exposed children, the cumulative risk was also related to delinquent and internalizing behaviors. We note that the term *delinquent* is not usually used with early elementary school-aged children and that the children's scores on the TRF delinquency scale were low. The principal TRF delinquent behaviors endorsed by the teachers were breaking school rules and lying or cheating. The data show that the cumulative risk was related to increased incidence of these behaviors at 7.5 years of age among the children exposed to substances prenatally.

It is of interest that the cognitive and behavioral endpoints examined in this study were influenced by different sets of risk factors. Whereas SES, family size, parental substance abuse, and parental violence were the principal risk factors for FSIQ, having an emotionally responsive primary caregiver and a warm emotional climate at home was the most consistent predictor of lower levels of delinquent and aggressive behaviors at school across groups. The presence of the biological father in the home also appears to be associated with reduced delinquent behavior. The inverse relation between maternal depression and externalizing behaviors in the nonexposed group, which was not predicted, suggests that family temperament or modeling parental behaviors may influence whether the child's psychosocial problems will be expressed in the form of delinquent or aggressive behaviors. The identification of the parent–child interaction as the most important socioenvironmental predictor of externalizing behavior problems is consistent with data from another African American sample (Jones et al., 2002). Just as much emphasis has been placed on the importance of intellectual stimulation and emotional support for cognitive development in middle-class Caucasian samples, quality of parenting also appears to be most important for socioemotional development and behavioral outcomes for inner-city African American children.

In the Isle of Wight study, Rutter (1979) found that the relation of the number of risk factors to which the child was exposed to developmental outcome was nonlinear, with a threshold of four or more. Whereas most children are relatively unaffected by three or fewer socioenvironmental risk factors, Rutter's data indicate that exposure to multiple risks interacts so that their effect on development becomes synergistic. It is of particular interest that, similar to Caucasian middle-class cohorts, four or more risks emerged as a threshold in our dose–response analyses of the African American children in our cohort who were not exposed prenatally to alcohol or cocaine for three important outcomes—VIQ, delinquent behaviors, and internalizing behaviors.

The multiple regression and correlational data in Table 5 did not indicate increased vulnerability to socioenvironmental risk in the prenatally exposed group; the multiple *R*s for the regressions and correlation coefficients for the cumulative risk index were generally slightly higher in the nonexposed group. Because the coefficients of variation for the risk factors were similar for the two groups (Table 2), this difference in effect size cannot be attributed to restriction of range in the exposed group. By contrast, the data from the dose – response analyses indicate that the dampening of the effect of exposure to three or fewer risk factors in the nonexposed group was not found in the prenatally exposed group. For most endpoints, each additional risk factor was associated with poorer cognitive performance and increased behavior problems in a linear fashion among the substance-exposed children. Moreover, for the one endpoint where a threshold was observed—delinquent behavior—behavior problems increased at a lower threshold (three or less risk factors) than among the nonexposed children. Another indication of increased vulnerability is the finding that cumulative risk was associated with increased internalizing behavior problems only for the prenatally exposed children. In addition, the data in Table 3 suggest that multiple risk factors are more likely to occur together in the families of the substance-exposed children. Although parental substance abuse, life stress, and maternal depression are moderately intercorrelated in both groups, for the prenatally exposed children quality of parenting, the principal protective factor for externalizing behavior problems, was also more correlated with SES, caregiver depression, and substance abuse, increasing the likelihood of the co-occurrence of multiple risks for the children in that group.

In interpreting the findings, certain limitations of the study should be noted. Because the focus of the study was on economically disadvantaged, inner-city African American families, one must be cautious about generalizing from these findings to other populations, who may have greater access to community or other forms of social support. Because prenatal substance exposure is associated with prematurity and reduced birth weight, the exclusion of very low birth weight and severe preterm deliveries may have reduced the number of more severely affected children in the cohort. In addition, although moderating effects of gender on externalizing behaviors have been reported in children prenatally exposed to alcohol and cocaine (Nordstrom-Bailey et al., 2005), we did not examine the interaction between gender and prenatal substance exposure in this study because the cell sizes in such analyses would have been unduly small. Nordstrom-Bailey et al.'s (2005) data also suggest that prenatal alcohol exposure may moderate prenatal cocaine effects on teacher-reported child externalizing behavior. We did not examine the effects of prenatal alcohol and cocaine exposure separately, however, to optimize cell sample sizes and because our hypothesis regarding fetal patterning focused on the impact of substance exposure generally on vulnerability to socioenvironmental risk.

Despite these limitations, these data lend support to findings of synergistic effects among multiple socioenvironmental risk factors reported in previous studies and provide evidence that four or more risks may be an important threshold for non-substance-exposed, inner-city African American children as well as the Caucasian children examined in previous studies. These data also provide support for the prediction from the fetal patterning of disease model that prenatal exposure to alcohol and/or cocaine can increase vulnerability to socioenvironmental risk. Based on the cumulative risk index and dose – response data, vulnerability to the adverse effects of socioenvironmental risk appears to be enhanced particularly on two endpoints—delinquent and internalizing behaviors.

References

Achenbach, T. M. (1991). *Integrative guide for the 1991 CBCL/4-18, YSR and TRF profiles*. Burlington: University of Vermont, Department of Psychiatry.

Alessandri, S. M., & Bendersky, M. L. (1998). Cognitive functioning in 8 to 18 month old drug-exposed infants. *Developmental Psychology, 34*, 565 – 573.

Attar, B. K., Guerra, N. G., & Tolan, P. H. (1994). Neighborhood disadvantage, stressful life events, and adjustment in urban elementary-school children. *Journal of Clinical Child Psychology, 23*, 391 – 400.

Beck, A. T., & Steer, R. A. (1987). *BDI, Beck Depression Inventory: Manual*. San Antonio, TX: Psychological Corporation.

Bendersky, M., Bennett, D., & Lewis, M. (2006). Aggression at age 5 as a function of prenatal exposure to cocaine, gender, and environmental risk. *Journal of Pediatric Psychology, 31*, 71 – 84.

Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology, 53*, 371 – 399.

Burchinal, M. R., Roberts, J. E., Hooper, S., & Zeisel, S. A. (2000). Cumulative risk and early cognitive development: A comparison of statistical risk models. *Developmental Psychology, 36*, 793 – 807.

Burden, M. J., Jacobson, S. W., & Jacobson, J. L. (2005). The relation of prenatal alcohol exposure to cognitive processing speed and efficiency in childhood. *Alcoholism: Clinical and Experimental Research, 29*, 1473 – 1483.

Burden, M. J., Jacobson, S. W., Sokol, R. J., & Jacobson, J. L. (2005). Effects of prenatal alcohol exposure on attention and working memory at 7.5 years of age. *Alcoholism: Clinical and Experimental Research, 29*, 443–452.

Caldwell, B., & Bradley, R. (1984). *Home Observation for Measurement of the Environment*. Little Rock: University of Arkansas at Little Rock.

Carmichael, O. H., Feldman, J. J., Streissguth, A. P., Sampson, P. D., & Bookstein, F. L. (1998). Neuropsychological deficits in adolescents with fetal alcohol syndrome: Clinical findings. *Alcoholism: Clinical and Experimental Research, 22*, 1998–2012.

Carta, J. J., Atwater, J. B., Greenwood, C. R., McConnell, S. R., McEvoy, M. A., & Williams, R. (2001). Effects of cumulative prenatal substance exposure and environmental risks on children's developmental trajectories. *Journal of Clinical Child Psychology, 30*, 327–337.

Coles, C. D., & Black, M. M. (2006). Introduction to the special issue: Impact of prenatal substance exposure on children's health, development, school performance, and risk behavior. *Journal of Pediatric Psychology, 31*, 1–4.

Compas, B. E., Howell, D. C., Phares, V., Williams, R. A., & Giunta, C. T. (1989). Risk factors for emotional/behavioral problems in young adolescents: A prospective analysis of adolescent and parental stress and symptoms. *Journal of Consulting and Clinical Psychology, 53*, 732–740.

Deater-Deckard, K., Dodge, K. A., Bates, J. E., & Pettit, G. S. (1998). Multiple risk factors in the development of externalizing behavior problems: Group and individual differences. *Development and Psychopathology, 10*, 469–493.

Derogatis, L. R. (1992). *SCL-90-R: Administration, scoring & procedures manual-II*. Baltimore: Clinical Psychometric Research.

Dodge, K. A. (1990). Developmental psychopathology in children of depressed mothers. *Developmental Psychology, 26*, 3–6.

Dunn, L. M., & Dunn, L. M. (1981). *PPVT manual for Forms L and M*. Circle Pines, MN: American Guidance Service.

Escalona, S. K. (1982). Babies at double hazard: Early development of infants at biologic and social risk. *Pediatrics, 70*, 670–676.

Famy, C., Streissguth, A. P., & Unis, A. S. (1998). Mental illness in adults with fetal alcohol syndrome or fetal alcohol effects. *American Journal of Psychiatry, 155*, 552–554.

Frank, D. A., Jacobs, R. R., Beeghly, M., Augustyn, M., Bellinger, D., Cabral, H., et al. (2002). Level of prenatal cocaine exposure and scores on the Bayley Scales of Infant Development: Modifying effects of caregiver, early intervention, and birth weight. *Pediatrics, 110*, 1143–1152.

Frank, D. A., & McCarten, K. M. (1999). Level of in utero cocaine exposure and neonatal ultrasound findings. *Pediatrics, 104*, 1101–1105.

Garmezy, N. (1985). Stress-resistant children: The church for protective factors. In J. Stevenson (Ed.), *Recent research in developmental psychopathology* (pp. 213–233). Oxford, England: Pergamon.

Gerard, J. M., & Buehler, C. (2004). Cumulative environmental risk and youth maladjustment: The role of youth attributes. *Child Development, 75*, 1832–1849.

Gluckman, P. D., & Hanson, M. A. (2004). Living with the past: Evolution, development and patterns of disease. *Science, 305*, 1733–1736.

Goldschmidt, L., Richardson, G. A., Stoffer, D. S., Geva, D., & Day, N. L. (1996). Prenatal alcohol exposure and academic achievement at age six: A nonlinear fit. *Alcoholism: Clinical and Experimental Research, 20*, 763–770.

Holden, G. W., & Ritchie, K. L. (1991). Linking extreme marital discord, child rearing, and child behavior problems: Evidence from battered women. *Child Development, 62*, 311–327.

Hollingshead, A. B. (1975). *Four factor index of social status*. Unpublished manuscript, Yale University.

Holmes, T. H., & Rahe, R. H. (1967). The Social Readjustment Rating Scale. *Journal of Psychosomatic Research, 11*, 213–218.

Hoyme, H. E., May, P. A., Kalberg, W. O., Kodituwakku, P., Gossage, J. P., Trujillo, P. M., et al. (2005). A practical clinical approach to diagnosis of fetal alcohol spectrum disorders: Clarification of the 1996 Institute of Medicine criteria. *Pediatrics, 115*, 39–47.

Jacobson, J. L., & Jacobson, S. W. (1994). Prenatal alcohol exposure and neurobehavioral development: Where is the threshold? *Alcohol Health and Research World, 18*, 30–36.

Jacobson, J. L., Jacobson, S. W., Sokol, R. J., & Ager, J. W. (1998). Relation of maternal age and pattern of pregnancy drinking to functionally significant cognitive deficit in infancy. *Alcoholism: Clinical and Experimental Research, 22*, 345–351.

Jacobson, S. W. (1998). Specificity of neurobehavioral outcomes associated with prenatal alcohol exposure. *Alcoholism: Clinical and Experimental Research, 22*, 313–320.

Jacobson, S. W., Carr, L. G., Croxford, J., Sokol, R. J., Li, T.-K., & Jacobson, J. L. (2006). Protective effects of the alcohol dehydrogenase-*ADH1B* allele in African American children exposed to alcohol during pregnancy. *Journal of Pediatrics, 148*, 30–37.

Jacobson, S. W., Chiodo, L. M., Jacobson, J. L., & Sokol, R. J. (2002). Validity of maternal report of alcohol, cocaine, and smoking during pregnancy in relation to infant neurobehavioral outcome. *Pediatrics, 109*, 815–825.

Jacobson, S. W., Dodge, N., Chiodo, L. M., Sokol, R. J., Dehaene, S., & Jacobson, J. L. (2003). Evidence for a specific effect of prenatal alcohol exposure on "number sense." *Alcoholism: Clinical and Experimental Research, 27*, 121A.

Jacobson, S. W., Jacobson, J. L., & Frye, K. F. (1991). Incidence and correlates of breast-feeding in socioeconomically disadvantaged women. *Pediatrics, 88*, 728–736.

Jacobson, S. W., Jacobson, J. L., Sokol, R. J., Chiodo, L. M., & Corobana, R. (2004). Maternal age, alcohol abuse history, and quality of parenting as moderators of the effects of prenatal alcohol exposure on 7.5-year intellectual function. *Alcoholism: Clinical and Experimental Research*, *28*, 1732–1745.

Jacobson, S. W., Jacobson, J. L., Sokol, R. J., Martier, S. S., & Ager, J. W. (1993). Prenatal alcohol exposure and infant information processing ability. *Child Development*, *64*, 1706–1721.

Jacobson, S. W., Jacobson, J. L., Sokol, R. J., Martier, S. S., & Chiodo, L. M. (1996). New evidence for neurobehavioral effects of in utero cocaine exposure. *Journal of Pediatrics*, *129*, 581–590.

Jones, D. J., Forehand, R., Brody, G., & Armistead, L. (2002). Psychosocial adjustment of African American children in single-mother families: A test of three risk models. *Journal of Marriage and Family*, *64*, 105–115.

Kim, S., & Brody, G. H. (2005). Longitudinal pathways to psychological adjustment among Black youth living in single-parent households. *Journal of Family Psychology*, *19*, 305–313.

Krishnakumar, A., & Black, M. M. (2002). Longitudinal predictors of competence among African American children: The role of distal and proximal risk factors. *Applied Developmental Psychology*, *23*, 237–266.

Landry, S. H., Smith, K. E., Swank, P. R., Assel, M. A., & Vellet, S. (2001). Does early responsive parenting have a special importance for children's development or is consistency across early childhood necessary? *Developmental Psychology*, *37*, 387–403.

Levy-Shiff, R., Einat, G., Mogilner, M. B., Lerman, M., & Krikler, R. (1994). Biological and environmental correlates of developmental outcome of prematurely born infants in early adolescence. *Journal of Pediatric Psychology*, *19*, 63–78.

Luster, T., & McAdoo, H. P. (1994). Factors related to the achievement and adjustment of young African American children. *Child Development*, *65*, 1080–1094.

Mattson, S. N., Riley, E. P., Gramling, L., Delis, D. C., & Jones, K. L. (1997). Heavy prenatal alcohol exposure with or without physical features of fetal alcohol syndrome leads to IQ deficits. *Journal of Pediatrics*, *131*, 718–721.

Mayes, L. C. (2002). A behavioral teratogenic model of the impact of prenatal cocaine exposure on arousal regulatory systems. *Neurotoxicology and Teratology*, *24*, 385–395.

Mayes, L. C., Molfese, D. L., Key, A. P. F., & Hunter, N. C. (2005). Event-related potentials in cocaine-exposed children during a Stroop task. *Neurotoxicology and Teratology*, *27*, 797–813.

McCloskey, L. A., Figueredo, A. J., & Koss, M. P. (1995). The effects of systemic family violence on children's mental health. *Child Development*, *66*, 1239–1261.

McDonald, R. J., Devan, B. D., & Hong, N. S. (2004). Multiple memory systems: The power of interactions. *Neurobiology of Learning and Memory*, *82*, 333–346.

McLoyd, V. C. (1990). The impact of economic hardship on Black families and children: Psychological distress, parenting, and socioemotional development. *Child Development*, *61*, 311–346.

Moss, H., Vanyukov, M., Majumder, P., Kirisci, L., & Tarter, R. E. (1995). Prepubertal sons of substance abusers: Influences of parental and familial substance abuse on behavioral disposition, IQ, and school achievement. *Addictive Behaviors*, *20*, 345–358.

Noble, K. G., McCandliss, B. D., & Farah, M. J. (2007). Socioeconomic gradients predict individual differences in neurocognitive abilities. *Developmental Science*, *10*, 464–480.

Nordstrom-Bailey, B., Sood, B. G., Sokol, R. J., Ager, J., Janisse, J., Hannigan, J. H., et al. (2005). Gender and alcohol moderate prenatal cocaine effects on teacher-report of child behavior. *Neurotoxicology and Teratology*, *27*, 181–189.

Nunes, E., Weissman, M. Goldstein, R., McAvay, G., Seracini, A. M., Verdeli, H., et al. (1998). Psychopathology in children of parents with opiate dependence and/or major depression. *Journal of American Academy of Child and Adolescent Psychiatry*, *37*, 1142–1151.

O'Connor, M. J., & Paley, B. (2006). The relationship of prenatal alcohol exposure and the postnatal environment to child depressive symptoms. *Journal of Pediatric Psychology*, *31*, 50–64.

O'Connor, M. J., Shah, B., Whaley, S., Cronin, P., Gunderson, B., & Graham, J. (2002). Psychiatric illness in a clinical sample of children with prenatal alcohol exposure. *American Journal of Drug and Alcohol Abuse*, *28*, 743–754.

Parker, S., Greer, S., & Zuckerman, B. (1988). Double jeopardy: The impact of poverty on early child development. *Pediatric Clinics of North America*, *35*, 1227–1240.

Puttler, L. I., Zucker, R. A., Fitzgerald, H. E., & Bingham, C. R. (1998). Behavioral outcomes among children of alcoholics during the early and middle childhood years: Familial subtype variations. *Alcoholism: Clinical and Experimental Research*, *22*, 1972.

Rutter, M. (1979). Protective factors in children's responses to stress and disadvantage. In M. W. Kent & J. E. Rolf (Eds.), *Primary prevention of psychopathology: Social competence in children* (pp. 49–74). Hanover, NH: University Press of New England.

Sameroff, A., Seifer, R., Baldwin, A., & Baldwin, C. (1993). Stability of intelligence from preschool to adolescence: The influence of social and family risk factors. *Child Development*, *64*, 80–97.

Selzer, M. L. (1971). The Michigan Alcoholism Screening Test: The quest for a new diagnostic instrument. *American Journal of Psychiatry*, *127*, 1653–1658.

Singer, L. T., Arendt, R., Minnes, S., Farkas, K., Salvator, A., Kirchner, H. L., et al. (2002). Cognitive and motor outcomes of cocaine-exposed infants. *Journal of American Medical Association*, *287*, 1952–1960.

Singer, L. T., Arendt, R., Minnes, S., Salvator, A., Siegel, A. C., & Lewis, B. A. (2001). Developing language skills of cocaine-exposed infants. *Pediatrics*, *107*, 1057–1064.

Spohr, H. L., & Willms, J. (1993). Prenatal alcohol exposure and long-term developmental consequences. *Lancet*, *341*, 907–910.

Steinhausen, H. C., Willms, J., & Spohr, H. L. (1993). Long-term psychopathological and cognitive outcome of children with fetal alcohol syndrome. *Journal of the American Academy of Child and Adolescent Psychiatry*, *32*, 990–994.

Stratton, K., Howe, C., & Battaglia, F. (1996). *Fetal alcohol syndrome: Diagnosis, epidemiology, prevention, and treatment*. Washington, DC: National Academy Press.

Straus, M. A. (1979). Measuring intrafamily conflict and violence: The Conflict Tactics (CT) Scales. *Journal of Marriage and the Family, 41*, 75–88.

Streissguth, A. P., Aase, J. M., Clarren, S. K., Randels, S. P., LaDue, R. A., & Smith, D. F. (1991). Fetal alcohol syndrome in adolescents and adults. *Journal of the American Medical Association, 265*, 1961–1967.

Streissguth, A. P., Barr, H. M., & Sampson, P. D. (1990). Moderate prenatal alcohol exposure: Effects on child IQ and learning problems at age 7 1/2 years. *Alcoholism: Clinical and Experimental Research, 14*, 662–669.

Tolan, P. H. (1988). Socioeconomic, family, and social stress correlates of adolescent antisocial and delinquent behavior. *Journal of Abnormal Child Psychology, 16*, 317–331.

Wadhwa, P. D. (2005). Psychoneuroendocrine processes in human pregnancy influence fetal development and health. *Psychoneuroendocrinology, 30*, 724–743.

Wasserman, G. A., Miller, L. S., Pinner, E., & Jaramillo, B. (1996). Parenting predictors of early conduct problems in urban, high-risk boys. *Journal of American Academy of Child and Adolescent Psychiatry, 35*, 1227–1236.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children* (3rd ed.). San Antonio, TX: Psychological Corporation.

Weinberg, N. Z. (1997). Cognitive and behavioral deficits associated with parental alcohol use. *Journal of American Academy of Child and Adolescent Psychiatry, 34*, 779–785.

Werner, E. E., & Smith, R. S. (1982). *Vulnerable but invincible: A longitudinal study of resilient children and youth*. New York: McGraw-Hill.

Wilens, T. E., Biederman, J., Bredin, E., Hahesy, A. L., Abrantes, A., Neft, D., et al. (2002). A family study of the high-risk children of opioid- and alcohol-dependent parents. *American Journal on Addictions, 11*, 41–51.

Yumoto, C., Corobana, R., Jacobson, S. W., Chiodo, L., Sokol, R. J., & Jacobson, J. L. (2002). The influence of father's presence in the home on alcohol-related aggression and delinquency in childhood. *Alcoholism: Clinical and Experimental Research, 26*, 92A.

Child Development, November/December 2008, Volume 79, Number 6, Pages 1777 – 1791

The Social Ecology of Adolescent Alcohol Misuse

Susan T. Ennett
The University of North Carolina at Chapel Hill

Vangie A. Foshee, Karl E. Bauman, Andrea Hussong, Li Cai, and Heathe Luz McNaughton Reyes
The University of North Carolina at Chapel Hill

Robert Faris
University of California at Davis

John Hipp
University of California at Irvine

Robert DuRant
Winston-Salem, North Carolina

A conceptual framework based on social ecology, social learning, and social control theories guided identification of social contexts, contextual attributes, and joint effects that contribute to development of adolescent alcohol misuse. Modeling of alcohol use, suggested by social learning theory, and indicators of the social bond, suggested by social control theory, were examined in the family, peer, school, and neighborhood contexts. Interactions between alcohol modeling and social bond indicators were tested within and between contexts. Data were from a longitudinal study of 6,544 students, 1,663 of their parents, and the U.S. Census. All contexts were uniquely implicated in development of alcohol misuse from ages 11 through 17 years, and most alcohol modeling effects were contingent on attributes of social bonds.

A social ecological perspective suggests that multiple social contexts and the interdependencies among contexts must be considered in explaining development of adolescent problem behaviors, such as alcohol misuse. Although adolescent alcohol use and misuse have been recognized as shaped by social contexts and processes (e.g., Brook, Nomura, & Cohen, 1989; Perry, Kelder, & Komro, 1993; Petraitis, Flay, & Miller, 1995), research on higher level contexts, such as schools and neighborhoods, is less common than research on peer and family contexts. Comprehensive examination of the multiple contexts comprising the social ecology of adolescent alcohol use is even more limited. Explanation for the lack of social ecological research may rest in the magnitude of conceptual and analytic demands when multiple social contexts that are themselves multidimensional are jointly considered.

We draw on theories of ecology of human development (Bronfenbrenner, 1977, 1979), social learning (Akers, Krohn, Lanza-Kaduce, & Radosevich, 1979; Bandura, 1977; Petraitis et al., 1995), and social control (Hirschi, 1969; Petraitis et al., 1995) to examine development of alcohol misuse in adolescence by specifying a set of social contexts, attributes of those contexts, and interrelations among those attributes to examine. Our purpose was to use these theories to identify a parsimonious set of contextual attributes, while broadly considering the multiple social contexts in which adolescents' lives are embedded, that in interaction with each other could meaningfully explain development of adolescent alcohol misuse.

Bronfenbrenner's Ecology of Human Development Theory

Bronfenbrenner's theory provides the overarching conceptual framework by establishing a developmental perspective; defining the social contexts for investigation; specifying the need for an inclusive, multidimensional view of these social contexts; and suggesting relationships of the contexts to each other and the

---

This research was supported by a grant from the National Institute on Drug Abuse (R01 DA13459). We thank Kate Karriker-Jaffe for geocoding assistance and James Moody for consultation on social network analysis. In addition, we thank the students, parents, and staff in the schools that participated in the study.

Correspondence concerning this article should be addressed to Susan T. Ennett, Department of Health Behavior and Health Education, The University of North Carolina at Chapel Hill, Campus Box 7440, Chapel Hill, NC 27599. Electronic mail may be sent to sennett@email.unc.edu.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7906-0017

developing adolescent. The central precepts of the ecology of human development are that human development takes place within a set of nested and changing environments and that complex relations within and between those nested environments shape behaviors. Bronfenbrenner suggests an approach to examining interrelations among social contexts by positing that "in ecological research, the principal main effects are likely to be interactions" (Bronfenbrenner, 1977, p. 518).

Bronfenbrenner defines person–environment interrelations in terms of micro-, meso-, exo-, and macrosystems. The first three systems are relevant to the current investigation because they define social contexts. Microsystems are the principal and immediate socialization contexts in which human development takes place: the family, peer, and school contexts for adolescents. An exosystem comprises more remote social environments, such as neighborhoods, in which microsystems are embedded. The concept of mesosystems is that processes operating in different contexts are not independent of each other. According to Bronfenbrenner, a mesosystem is formed from the interrelations among microsystems; as an example, processes in the family may impinge on relationships in the peer group (Bogenschneider, Wu, Raffaelli, & Tsay, 1998).

Theoretical Perspectives Relevant to Specifying Contextual Attributes

Despite the conceptual utility of the ecology of human development, Bronfenbrenner does not provide specificity about the particular attributes of social contexts to measure. Social learning and social control theories are helpful here because both theories situate causes of alcohol use and other problem behaviors in the social environment. Moreover, they are the dominant theoretical perspectives in research on adolescent alcohol and other substance use and have substantial empirical support (Petraitis et al., 1995).

Social learning theory emphasizes exposure to alcohol-using role models and social control theory focuses on the constraining function of social bonds. Social learning theory posits that adolescent alcohol use is a learned behavior acquired through social interactions and reinforcement and the resulting ratio of exposure to prosocial versus antisocial definitions for drinking (Petraitis et al., 1995). The ubiquitous positive association between adolescent drinking and their friends' drinking (e.g., Bauman & Ennett, 1996) is typically understood as evidence of social learning, as is the relationship between parental alcohol use and drinking in their children (e.g., Richter & Richter, 2001).

Social control theory posits that a tendency toward deviance is universally shared but manifested only when the bond between an individual and society is weakened (Hirschi, 1969; Petraitis et al., 1995). The social bond may be weakened when conventional attachments are attenuated as when parent–adolescent closeness is low or parents do not exercise supervision of their adolescents (e.g., Hops, Duncan, Duncan, & Stoolmiller, 1996; Simons-Morton & Chen, 2005), and in the face of stress, such as in high-conflict families (e.g., Aseltine & Gore, 2000; Webb & Baer, 1995). At a higher level, and often understood from the perspective of macrolevel social control theory, adolescent deviance may be due to attenuated bonds in socially disorganized neighborhoods (Elliott et al., 1996; Sampson, Raudenbush, & Earls, 1997).

Conceptual Framework

Figure 1 presents a schematic representation of the conceptual framework guiding the study. Based on the ecology of human development, four social contexts are posited as potentially relevant to development of adolescent alcohol misuse—family, peers, school (microsystems), and neighborhood (exosystem). The arrows between the family, peer, and school contexts indicate interrelations among these microsystem contexts (mesosystems), thus between family and peers, family and school, and peers and school.

Consistent with Bronfenbrenner's injunction to measure the complexity of social contexts, each context is characterized by a set of four variables. The variables are derived from social learning or social control theory, with each variable tailored to the particular context. Modeling of alcohol use by others (family members, friends, schoolmates, neighbors) is measured as the central construct from social learning theory. From social control theory, closeness to others, social regulation, and stress are measured as indicators of social bonds pertinent to constraining deviant behavior. As an example of the tailoring of these social control variables to the context, closeness is operationalized as parent–child closeness in the family context, feelings of closeness among mutual friends in the peer context, overall level of school bonding among all students in the school context, and neighborhood bonding among neighborhood residents.

All four social contexts are expected to uniquely contribute to adolescent alcohol misuse development. In addition, following from Bronfenbrenner's prediction of the primacy of interactive effects, the social learning and social control variables characterizing the contexts are expected to have interactive effects within the four contexts and between the three

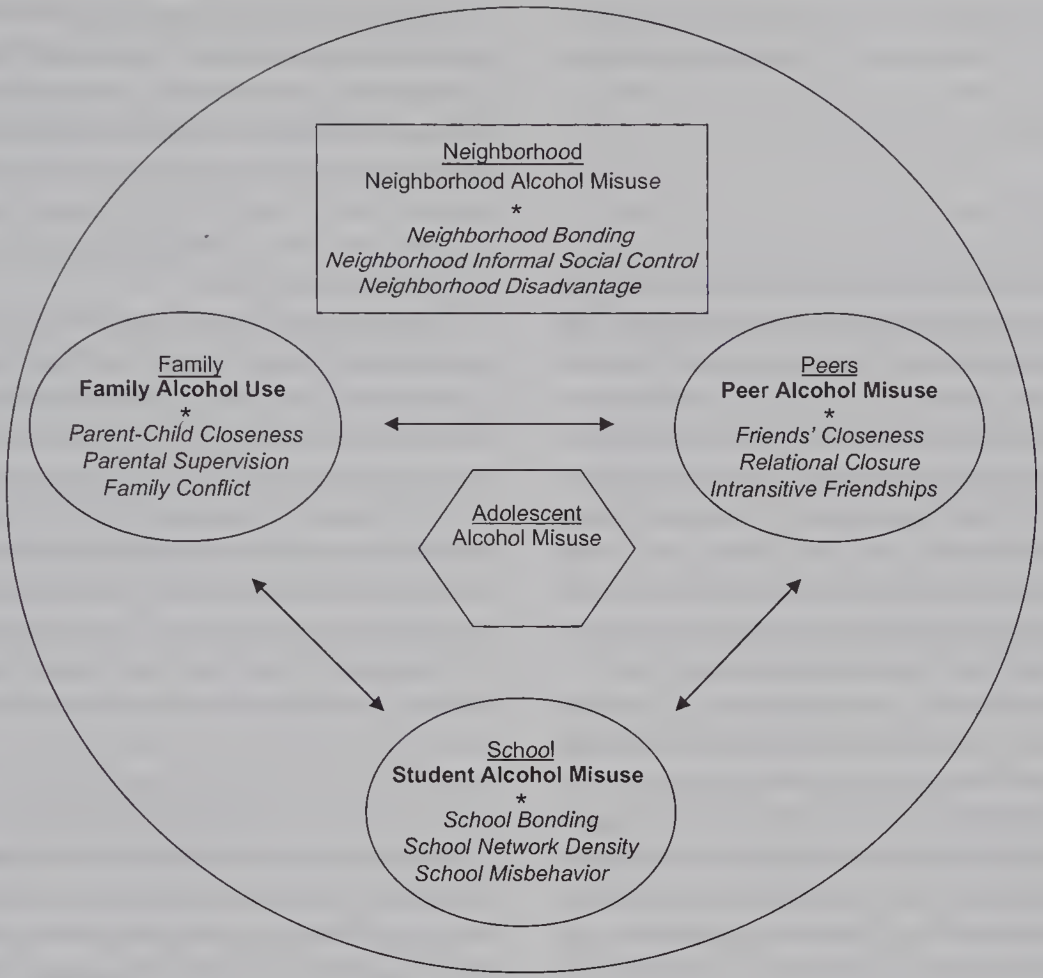

Neighborhood
Neighborhood Alcohol Misuse
*
Neighborhood Bonding
Neighborhood Informal Social Control
Neighborhood Disadvantage
Family
Family Alcohol Use
*
Parent-Child Closeness
Parental Supervision
Family Conflict
Peers
Peer Alcohol Misuse
*
Friends' Closeness
Relational Closure
Intransitive Friendships
Adolescent
Alcohol Misuse
School
Student Alcohol Misuse
*
School Bonding
School Network Density
School Misbehavior

Figure 1. Conceptual framework of the social context of adolescent alcohol misuse based on social ecology, social learning, and social control theories.

Note. Family, peer, school, and neighborhood contexts and their relationships to each other are suggested by the ecology of human development; ovals depict microsystems, arrows depict mesosystems, and the rectangle depicts an exosystem. Characteristics of contexts in bold are derived from social learning theory (modeling of alcohol use) and characteristics in italics are derived from social control theory (closeness, social regulation, stress).

microsystem contexts. Specifically, alcohol modeling effects are expected to vary depending on attributes of the social bonds.

Empirical Support for Interactive Effects Between Social Learning and Social Control Constructs Between and Within Social Contexts

Studies of adolescent alcohol or other substance use in which interactions among social learning and social control constructs were examined are surprisingly limited, whether considering studies that simultaneously considered two or more social contexts or that examined only a single social context. When two or more contexts have been included, family–peer relations have been most commonly studied and mediating effects between social learning and social control constructs examined more often than moderating effects. These studies provide evidence that family attributes such as drug models, parent–adolescent attachment quality, parenting style, and family structure buffer negative peer effects (Brook, Brook, Gordon, Whiteman, & Cohen, 1990; Brook, Whiteman, Gordon, Nomura, & Brook, 1986 A. D. Farrell & White, 1998; Marshal & Chassin, 2000; Mason, Cauce, Gonzales, & Hiraga, 1994).

Several studies have examined interactive effects of social learning and social control constructs on adolescent alcohol use in the family context and, less often, the peer context. Results are mixed in supporting conditional relationships when effects of the parent–child relationship and parent alcohol use were examined. Some studies found evidence for no interactions between parent–adolescent attachment quality and parental alcohol use (Andrews, Hops, & Duncan, 1997; Kandel & Andrews, 1987), whereas others reported amplifying effects of lower levels of attachment quality on parental alcohol use and/or buffering effects of higher levels of attachment quality (Andrews et al., 1997; M. P. Farrell, Barnes, & Banerjee, 1995; Urberg, Goldstein, & Toro, 2005; Zhang, Welte, & Wieczorek, 1997). In the peer context, the relationship between peer and adolescent alcohol use has been shown to be stronger when peer support is high

(Urberg et al., 2005). Similarly, in research on adolescent delinquency, delinquency–peer associations may be stronger among closer friends (Giordano, Cernkovich, & Pugh, 1986; Haynie, 2001).

Study Hypotheses

We test the fundamental premises of the ecology of human development that attributes of multiple social contexts will contribute to development of adolescent alcohol misuse and that interrelations within and between contexts, tested through interactions between social learning and social control constructs, will be important effects. Specifically, we hypothesize that the effects of alcohol use by others will vary depending on closeness, social regulation, and stress, and that the nature of the interactions will vary depending on the context or contexts involved.

Within the family and neighborhood contexts, where adult norms against alcohol misuse are expected, we hypothesize that closeness and social regulation will buffer effects of family and neighbors' alcohol use. Within the peer and school contexts, where the reference is to interactions with other adolescents, closeness and social regulation are hypothesized to amplify effects of friends' and schoolmates' alcohol use. Consistent with these hypotheses, we expect that between-context interactions involving family closeness and social regulation will buffer exposure to alcohol use by friends and schoolmates, whereas peer closeness and social regulation will amplify exposure to alcohol use by schoolmates. Within and between all contexts, we hypothesize that stress will magnify the effect of alcohol use by others. In keeping with the developmental perspective of the ecological framework, we examine the contribution of these contextual effects to growth in adolescent alcohol misuse by modeling trajectories of alcohol misuse from age 11 to age 17 in a general population sample.

Method

Study Overview

The data were from a longitudinal investigation of intrapersonal and contextual factors that influence adolescent alcohol use and other problem behaviors. The study design included four components to enable contextual analyses: (a) a census of adolescents identified by school enrollment and surveyed in school every 6 months for a total of five assessments (Waves 1–5), (b) a simple random sample of parents of the adolescents surveyed annually by telephone for a total of three assessments (Waves 1, 3, and 5), (c) social network analysis of school networks based on friendship nominations collected in the school surveys, and (d) geocoding of all adolescent and parent addresses to allow linkage of the addresses to U.S. Census tracts and block groups. Data collection with adolescents and parents began in spring 2002 and ended in spring 2004.

All protocols were reviewed and approved by the institutional review boards at the University of North Carolina at Chapel Hill and Wake Forest University. The boards approved a waiver of written parental consent. Four weeks prior to data collection, parents were informed about the study through a letter sent both by first-class mail and home with the child. Parents could refuse their child's participation by returning a postage-paid form or by calling a toll-free number. Written adolescent assent was obtained in school at the time of data collection by trained data collectors.

Adolescent Sample and Data Collection

Adolescents enrolled in three public school systems in North Carolina were entered into the study when they were 6th, 7th, and 8th graders and completed the study as 8th, 9th, and 10th graders, respectively. The school systems included a total of eight middle schools, two comprehensive K–8 schools, six high schools, and three alternative schools with middle and high school grades. The high schools were not included until Wave 2 when the first 8th graders transitioned from middle school. Data collection was timed to coincide with the beginning and end of the school year. At each assessment, all enrolled students at the targeted grade levels, except for those in self-contained classrooms for exceptional children and those with limited English language reading skills, were eligible for the study. Thus, students new to the study enrolled at each wave of data collection; approximately 15% of students were newly enrolled at the beginning of each school year and approximately 3% were newly enrolled midyear. Across the five assessments, the samples ranged in size from 5,220 (Wave 1) to 5,017 (Wave 5) adolescents, with 6,891 unique cases across all waves. Response rates at the five waves were 88.4%, 81.3%, 80.9%, 79.1%, and 76.0%, respectively. Approximately one third of students were in each of the three grade cohorts; at Wave 1 the distributions were 35.6%, 33.5%, and 30.9% in Grades 6, 7, and 8, respectively.

Trained data collectors administered questionnaires on at least two occasions at each school to reduce the effect of absenteeism on the response rate. Adolescents completed the questionnaire in approximately 1 hr. Names were not written on questionnaires, and questionnaires were placed in envelopes without names and sealed before returning them to the data collectors. Teachers stayed in classrooms to help maintain order but did not answer questions or walk around the classroom.

Parent Sample and Data Collection

A simple random sample of 1,663 parents of adolescents who completed the Wave 1 survey completed a 25-min telephone interview at Waves 1, 3, and 5. All parents completed the interview in English. Unlike for adolescents, new parents did not enter the study after Wave 1. By design, in the majority of cases (98.2%) the mother or mother surrogate was the parent interviewed. Response rates at the three waves were 79.8%, 82.5%, and 71.8%, respectively. The parent sample was designed to provide sufficient number of cases within each census geographic area to allow aggregate measures of neighborhood context from parent responses to questions about their neighborhood.

Based on self-reports, the mean age of parent respondents at Wave 1 was 40.50 years (*SD* = 7.23). Approximately 59% self-identified as White, 38% as Black, 1% as Hispanic, and 2% as another race/ethnicity. Approximately two thirds (64.9%) reported living in a two-parent household. For 45%, the highest education attained was high school graduate or less.

Social Network Analysis

Social network analysis was conducted on friendships reported by adolescents at each wave of data collection (Ennett et al., 2006). Data collectors provided each student a student directory that alphabetically listed all enrolled students along with a unique four-digit peer identification number for each student. Adolescents identified up to their five closest friends, starting with their best friend. Friends not in the directory were identified by "0000." Because most adolescent friendships are with adolescents in the same school and grade, we bounded social networks by school and grade, with the following exceptions. In high schools and alternative schools, networks were bounded by school only because classes and activities were not grade segregated and therefore cross-grade friendships were likely. In the two K–8 schools, networks also were bounded by school because of their small enrollments. We defined the peer context by the adolescent's set of close friends, including those nominated by the adolescent and those who nominated the adolescent as friend. The school context was defined by the school network to which the adolescent belonged and thus included a larger set of peers than close friends.

Geocoding

The addresses of adolescents and parents at each wave of data collection were sent to a commercial geocoding firm to be matched with U.S. Census tract and block group data in the 2000 decennial census. At Wave 1, 94.6% of the parent sample was matched to a street address and 5.4% was matched to a ZIP centroid; 94.8% of the adolescent sample was matched to a street address, 3.9% to the ZIP centroid, and less than 1% was not matched. The match rates were similar at other waves of data collection. The geocodes represented 158 census block groups and 56 tracts, and they covered all the block groups and tracts in the three-county area as well as some block groups and tracts in contiguous areas where some of the students and parents lived.

We defined the neighborhood context by the U.S. Census block group in which the adolescent lived the first time the adolescent was assessed. Neighborhood assignment was based on block group rather than tract because the smaller size more nearly approximated neighborhoods and comparison of intraclass correlation coefficients for neighborhood measures defined for block groups and census tracts showed greater homogeneity within block groups (data not shown).

Analysis Sample

The analysis sample included all adolescents who participated at any wave of data collection (*N* = 6,891) except for those missing birth date or outside the typical age range of 11–17 years for the grades studied (*n* = 51, 0.7%) or who could not be geocoded (*n* = 296, 4.3%), yielding a sample size of 6,544 (95.0%).

The mean self-reported age of adolescents at Wave 1 was 13.12 years (*SD* = 1.04). About half were male (51%) and the self-reported race/ethnicity distribution was 52% White, 37% Black, 4% Hispanic, and 7% other race/ethnicity. Averaged across all five waves of assessment, approximately 13% of adolescents reported living in other than a two-parent family and for 39% the highest education attained by either parent was reported by the adolescent to be high school or less.

Measures

Measures included adolescent alcohol misuse, social context variables suggested by social learning and social control theories, and demographic control variables. The alcohol and social context variables were collected at all five assessments, allowing construction of time-varying measures. With the exceptions of age and high school enrollment, the demographic variables were constructed as time invariant.

Alcohol misuse. We constructed a scale of alcohol misuse based on eight adolescent self-report items about recent alcohol use. Items measured problematic levels of use (e.g., had five or more drinks in a row, gotten drunk or very high from drinking alcoholic beverages) and negative consequences associated with use (e.g., gotten into a physical fight because of drinking, did something you later regretted because you had been drinking). Each item had five response categories ranging from 0 to 10 or more times in the past 3 months. Because of the skewed distribution of responses, as expected for a general population sample of young people, rather than constructing the scale by summing or averaging responses to the eight items, we used item response theory (IRT) to construct the scale (Thissen, Nelson, Rosa, & McLeod, 2001). Because IRT models require independent cases, we randomly selected one observation per respondent from all the repeated measures. We confirmed that a one-factor model fit the data well, using the weighted least squares estimator implemented in Mplus (Muthén & Muthén, 1998–2006). We used MULTILOG software to run both the graded response model and the nominal model in a comparison to determine which provided a better fit to the data (Thissen, Chen, & Bock, 2003). The nominal model provided a substantially better fit than the graded response model to the observed proportions in the lowest response category (i.e., zero times) for all items. In addition, even though the nominal model was unconstrained, it preserved the ordering of the response categories. Thus, we selected the nominal model. We then used the item parameter estimates to compute scale scores using the maximum a posteriori method (Thissen & Orlando, 2001). The resulting scale more nearly approximated a continuous distribution, with reduced skewness and kurtosis, compared with a summed scale.

Social context. We measured indicators of alcohol modeling, closeness, social regulation, and stress in each social context. Most measures were means of reduced sets of items from existing scales. Specific items were identified for survey inclusion through earlier psychometric analysis of data collected on full scales in a pilot study. Except where otherwise noted, items had either four or five response categories. Cronbach's alphas reported with the measures were averaged across the five waves.

Family context measures, based on adolescent report, were the number of current alcohol users in the family (alcohol modeling), parent–adolescent closeness (closeness), parental supervision (social regulation), and family conflict (stress). The number of drinkers was constructed from items about recent alcohol behavior asked separately for the mother, father, and siblings. Parent–adolescent closeness was measured by two sets of three questions asked separately about the mother and father, such as how close the adolescent felt toward the parent ($\alpha = .80$). Parental supervision was constructed from two sets of three items asked separately for the mother and father from the Authoritative Parenting Scale (Jackson, Henriksen, & Foshee, 1998). The scale measured parental supervisory behavior, such as whether the parent had rules that the adolescent must follow ($\alpha = .84$). Family conflict was measured by three items tapping fighting, throwing things, and sometimes hitting within the family ($\alpha = .85$; Bloom, 1985).

The peer measures were mean alcohol misuse among friends (alcohol modeling), mutual friends' closeness (closeness), relationship closure (social regulation), and adolescent membership in intransitive friendship triads (stress). Friends' alcohol use was based on the set of close friends and was the mean of their self-reported alcohol misuse. Peer closeness measured the mean closeness reported by mutual friends (i.e., friends who reciprocated the adolescent's nominations); thus, it measured how close these friends felt toward the adolescent. Relationship closure, based on adolescent reports, was the mean of three dichotomous items per nominated friend measuring whether the adolescent's parents had met the friend, the adolescent had met the friend's parents, and adolescent and friend's parents had met (Bearman & Moody, 2004). Higher values indicated greater parental knowledge of friends, suggesting both greater intimacy between adolescents and the possibility of more effective parental oversight and regulating influence on the peer context (Coleman, 1988). Peer stress was derived from the social network analysis (Moody, 2000, 2001) that identified intransitive triads to which adolescents belonged, where an intransitive triad was a set of three peers linked through friendship nominations such that a friend's friend was not also a friend of the adolescent. The variable was the proportion of intransitive triads among all the triads to which the adolescent belonged. Whereas transitive triads represented balanced, closed friendship circles, intransitive triads reflected unbalanced and potentially discordant relationships where an adolescent's friendships were not overlapping and thus carried the potential for stress (Bearman & Moody, 2004).

School context measures were based on data from all students in the school network. Measures were mean alcohol misuse in the network (alcohol modeling), school bonding (closeness), relative density of the school network (social regulation), and prevalence of school misbehavior (stress). School bonding was the mean of adolescents' agreement with three items, such as "students at this school are willing to go out of their way to help someone," from a scale measuring caring school communities (α = .84) (Battistich & Hom, 1997). School network relative density, identified through the social network analysis, was the proportion of possible friendship ties present in the school network; the measure adjusted for the fixed number of friend nominations (Moody, 2000, 2001). Higher values indicated greater presence of friendships linking together adolescents in the school and thus more closely knit school communities with presumed greater potential for regulating behavioral norms. School misbehavior was measured with five items from the Problem Behavior Frequency Scale (A. D. Farrell, King, White, & Valois, 2000) that tapped frequency of school-related problem behaviors, such as skipping school and threatening a teacher. Responses were averaged across all students in the school network. Schools with high levels of student misbehavior were assumed to be a stressful environment for adolescents.

We measured neighborhood context attributes, with neighborhoods defined by census block group, based on data from parents residing in each block group, adolescent residents, and the U.S. Census. Because parents were interviewed at Waves 1, 3, and 5 only and because we operationalized time-varying measures for all social context variables, we used the same parent-based neighborhood measures at Wave 2 as at Wave 1, and at Wave 4 as at Wave 3. The measures were the mean alcohol misuse of all adolescents living in the neighborhood (alcohol modeling), mean neighborhood bonding (closeness), mean informal neighborhood social control (social regulation), and mean neighborhood disadvantage (stress). All measures were grand mean centered.

Neighborhood bonding was based on parents' agreement with three items that measured bonding and trust among neighbors (α = .90; Parker et al., 2001). Neighborhood informal social control was constructed from six items that assessed willingness of neighbors to step in and do something when teens were misbehaving, such as damaging property or hanging out and smoking marijuana (α = .91; Sampson et al., 1997). Neighborhood disadvantage was based on parent agreements' with six items regarding physical and social disorder in the neighborhood, such as there being a lot of crime in the neighborhood (Ross & Jong, 2000), and five U.S. Census measures of socioeconomic disadvantage for the block group, such as percentage below the poverty line and percentage of households without a high school degree (α = .84).

Demographics. All measures were based on adolescent self-reports except for high school enrollment, which was derived from school records. Age was measured continuously based on date of birth. Sex was coded so the reference group was female. Race/ethnicity was based on the adolescent's modal response across all waves of assessment and dummy coded to include White (reference group), Black, Hispanic, and other race/ethnicity. Family structure was coded as two parents in the home (reference group) versus some other composition. Parent education measured the highest education attained by either parent and coded as high school graduate or less versus more than high school graduate (reference group). Because family structure and parent education could change over the course of the study, the two variables were coded as the average across the five waves.

Multiple Imputation

As already noted, students could enter the study at any of the five assessments and some students attrited at one or more waves. Of the 6,544 students in the analysis sample, 44.2% participated in all five waves, 18.3% participated in four waves, 15.2% in three waves, 10.9% in two waves, and 11.5% at only one wave. Adolescents missing at one or more waves compared with those continuously assessed once they entered the study were significantly more likely to be male, Black or of other race/ethnicity, live in other than a two-parent household, have parents with lower education, and have higher levels of alcohol use. Of the 1,663 parents, 66.8% completed all three interviews, 20.4% completed two interviews, and 12.8% completed only the first interview. Parents who completed one or two interviews compared with those who completed all three interviews were more likely to be Black, live in other than a two-parent household, and have lower education.

So that complete data sequences were available for all cases (Allison, 2002), we used PAN, a multilevel multiple imputation program appropriate for longitudinal data, to impute five sets of missing values (Schaefer, 2001). Variables included in the missingness equation were the demographics, alcohol misuse, and all social contextual variables at all five waves. We bounded the imputed values to the valid ranges of the data, and allowed all imputed dichotomous variables to range between 0 and 1 rather than rounding values (Allison, 2005).

Statistical Analysis

Because of the nestedness of our data and social contexts, such that repeated measures of alcohol misuse were nested within adolescents and adolescents were nested within neighborhoods and schools, we used a multilevel modeling approach (e.g., Raudenbush & Bryk, 2002). Specifically, we estimated three-level hierarchical growth models with time specified at Level 1, adolescents at Level 2, and neighborhood at Level 3. Four-level models cross-classified by neighborhood and school could not be estimated. We specified neighborhood rather than school at Level 3 because of the larger number of neighborhoods than schools, because neighborhoods were largely nested within schools (in that adolescents from several neighborhoods attended the same school) and because there was significant variance in adolescent alcohol misuse between neighborhoods (intraclass correlation coefficient [ICC] = .003, $p < .05$) but not schools (ICC = .001, $p = .18$).

The data were arranged in a cohort sequential design whereby data collected over approximately 2.5 years from three cohorts (i.e., cohorts enrolled at Grades 6, 7, and 8) were merged to allow accelerated trajectories of alcohol misuse to be modeled over approximately 6 years. We used age to measure the passage of time, thereby allowing change in alcohol misuse to be modeled from age 11 through age 17 (Mehta & West, 2000; Miyazaki & Raudenbush, 2000).

We report the analysis in stages beginning with estimation of the unconditional model to determine the random components and form of the trajectory, with an a priori expectation of a linear model. We then tested the cohort sequential design assumption of no differences across the three cohorts in the association between age and alcohol misuse by using a likelihood ratio test to compare the unconditional model to a model that added a variable measuring cohort and the interaction between cohort and age (Miyazaki & Raudenbush, 2000). A nonsignificant result would suggest the appropriateness of merging data across cohorts.

We next estimated a series of conditional models, using the concepts of micro-, exo-, and mesosystems to guide model building. All models controlled for the demographic variables and, due to multicollinearity, did not include the school context measure of stress. Model 1, the microsystem model, included the three sets of variables describing the family, peer, and school contexts. The variable sets included the main effects of the four attributes in each context and the two-way within-context interactions between alcohol modeling and closeness, social regulation, and stress. Model 2, the exosystem model, added to this microsystem model the set of neighborhood variables and their within-context interactions. Three mesosystem models then added to the exosystem model the two-way between-context interactions between alcohol modeling in one context, and closeness, social regulation, and stressful relations in the other context for the family and peer contexts (Model 3), family and school contexts (Model 4), and peer and school contexts (Model 5). We constructed the interactions such that, in Models 3 and 4, family characteristics moderated the relationships between adolescent alcohol misuse and peer/school alcohol modeling (e.g., family closeness moderated the relationship between adolescent and friends' alcohol misuse), and in Model 5, peer characteristics moderated the relationship between adolescent alcohol misuse and school alcohol modeling. We also included the between-context interaction between the two indicators of alcohol modeling because of the risk associated with accumulated exposure (Ary, Tildesley, Hops, & Andrews, 1993).

For each conditional model, we report the set of coefficients for the fixed effects of the social context attributes and the two-way interactions between context attributes. We also report the *F* statistic for testing the significance of the set of variables added to each successive model. Because the social context variables were time varying, a significant effect means that the relationship between the social context variable (or interaction between context variables) and adolescent alcohol misuse was significant on average over the ages examined. For ease of interpretation, we refer throughout to each contextual attribute by the construct name (e.g., peer stress) rather than the specific indicator (e.g., intransitive friendships triads). Significant interactions were probed by plotting the regression of alcohol misuse on alcohol user models at values of the moderator variable set at the mean and 1 *SD* above and below the mean (Curran, Bauer, & Willoughby, 2004).

All analyses were conducted using SAS version 9.1.3., using PROC MIXED and PROC MIANALYZE to estimate the unconditional and conditional mixed models (SAS, 2002–2003).

Results

Unconditional Model

A linear model with two random components, the individual and neighborhood intercepts, demonstrated significant individual (Z score = 42.06, $p < .0001$) and between-neighborhood (Z score = 2.96, $p < .001$) variation around the mean intercept for alcohol use centered at age 12. Linear models that included a random slope at the individual and neighborhood levels or at the individual level could not be estimated. The model also showed significant fixed effects such that the mean intercept for alcohol misuse was significantly different from zero ($B = -.02$, $SE = .01$, $p < .05$) and there was significant linear growth in the mean slope ($B = .10$, $SE = .00$, $p < .0001$) through age 17.

We compared this model to quadratic, unstructured, and spline models, each of which also included random intercepts at the individual and neighborhood levels. Based on the Bayesian information criterion (BIC) fit statistic, with smaller values by three points or more indicating better fit (Raftery, 1995), the linear model was favored over either the quadratic or the unstructured model (linear BIC = 45,570; quadratic BIC = 45,573; unstructured BIC = 45,615). Although the difference between the BICs for the linear and quadratic models was 3 points, the fixed quadratic effect was not significant and the likelihood ratio test comparing the two models was not significant. The spline model did not demonstrate sufficient improvement in model fit (BIC = 45,568) to be selected.

Test of Cohort Differences

The likelihood ratio test comparing linear models with and without the addition of a cohort variable and cohort by age interaction term was not statistically significant, indicating a lack of cohort effects and the appropriateness of the cohort sequential design.

Conditional Models

Demographic variables. We first examined the contributions of the demographic variables to adolescent alcohol misuse. The positive linear effect of age on alcohol misuse was significant after controlling for the other demographic variables ($B = .10$, $SE = .01$, $p < .0001$). High school enrollment, the other time-varying demographic characteristic, was significantly positively associated with alcohol misuse trajectories ($B = .05$, $SE = .01$, $p < .0001$). For the time-invariant variables, we examined the effects of the variables on the intercept and slope of the trajectories. Initial levels of alcohol misuse were higher only for youth whose parents had lower rather than higher educational levels ($B = .05$, $SE = .02$, $p < .05$). The rate of change in alcohol misuse was slower for Black youth compared with White youth ($B = -.06$, $SE = .01$, $p < .0001$) and faster for youth in families without two parents ($B = .04$, $SE = .01$, $p < .01$).

Microsystem model. Results of the microsystem, exosystem, and mesosystem models are shown in Table 1. For the microsystem model, after adjusting for the demographic variables and all other variables in the model, all the family context variables and the two-way interactions between family alcohol use modeling and family social bond indicators were significantly associated with alcohol misuse from ages 11 through 17. As expected, the indicators of family closeness and social regulation were negatively associated with adolescent alcohol misuse, whereas family stress and family alcohol use were positively associated with adolescent behavior. As hypothesized, family closeness ($p < .10$) and social regulation buffered the effect of family alcohol modeling on adolescent alcohol misuse, whereas family stress magnified the effect.

In the peer context, only the two-way interaction between peer social regulation and peer modeling of alcohol use was significant. In an unexpected finding, peer modeling of alcohol use was negatively associated with adolescent alcohol misuse, but greater social regulation buffered the negative effect. (The bivariate association between peer modeling of alcohol use and adolescent alcohol misuse was positive, as expected.) In the school context, alcohol modeling by all students was significantly positively associated with adolescent alcohol misuse; none of the social control variables were significantly associated with adolescent alcohol misuse.

Exosystem model. With the addition of the set of neighborhood variables to the microsystem model, neighbors' alcohol misuse had a significant positive effect on adolescent alcohol misuse. No other predictors or any of the interaction terms was significant. The effects of the family, peer, and school variables remained unchanged except that the interaction between peer stress and peer modeling of use became statistically significant. The negative (protective) effect of peer modeling was attenuated by higher stress.

Mesosystem models. For all three mesosystem models, the sets of Family × Peers, Family × School, and Peers × School interactions were significant additions to the exosystem model. In the Family × Peers and Family × School models, family social regulation buffered effects of both friends' and schoolmates'

Table 1
Social Context Predictors of Alcohol Misuse From Age 11 Through Age 17 (N = 6,544)

	Model 1	Model 2	Model 3	Model 4	Model 5
			Mesosystem		
	Microsystem	Exosystem	Family × Peers	Family × School	Peers × School
Social context	*B* (*SE*)	*B* (*SE*)	*B* (*SE*)	*B* (*SE*)	*B* (*SE*)
Family context					
Modeling	.12 (.02)****	.11 (.02)****	.07 (.02)***	.06 (.02)***	.11 (.02)****
Closeness	−.02 (.01)*	−.03 (.01)**	−.02 (.01)	−.03 (.01)†	−.03 (.01)**
Social regulation	−.04 (.01)****	−.03 (.01)****	−.01 (.01)	−.02 (.01)*	−.03 (.01)****
Stress	.03 (.01)****	.03 (.00)****	.01 (.01)	.01 (.01)*	.03 (.00)****
Closeness × Model	−.01 (.01)†	−.01 (.01)†	−.01 (.01)	−.00 (.01)	−.01 (.01)†
Regulation × Model	−.01 (.00)**	−.01 (.01)**	−.01 (.01)*	−.01 (.01)*	−.02 (.01)**
Stress × Model	.03 (.00) ****	.02 (.00)****	.02 (.00)****	.02 (.00) ****	.02 (.00) ****
Peer context					
Modeling	−.07 (.10)	−.49 (.10)****	−.38 (.17)*	−.47 (.10)***	−.57 (.13)***
Closeness	−.01 (.00)	−.00 (.01)	−.00 (.01)	−.00 (.01)	.00 (.01)
Social regulation	.01 (.01)	.01 (.01)	.01 (.01)	.01 (.01)	.00 (.01)
Stress	.02 (.05)	−.01 (.04)	−.00 (.04)	−.01 (.04)	−.03 (.05)
Closeness × Model	.01 (.02)	−.00 (.02)	−.00 (.02)	−.00 (.02)	.00 (.03)
Regulation × Model	.11 (.04)*	.12 (.04)*	.12 (.04)*	.12 (.04)*	.08 (.05)
Stress × Model	.45 (28)	.58 (.20)**	.53 (.20)*	.56 (.20)**	.34 (.25)
School context					
Modeling	.47 (.16)*	.52 (.15)**	.43 (.15)**	.03 (.20)	.13 (.18)
Closeness	−.00 (.02)	.00 (.02)	−.01 (.02)	−.01 (.02)	−.00 (.02)
Social regulation	.03 (.10)	.10 (.08)	.08 (.08)	.08 (.08)	.04 (.09)
Closeness × Model	−.04 (.10)	−.06 (.08)	−.02 (.08)	.02 (.08)	−.05 (.08)
Regulation × Model	−.03 (.34)	−.39 (.27)	−.32 (.27)	−.30 (.27)	−.21 (.29)
Neighborhood context					
Modeling		.67 (.03)****	.67 (.03)****	.67 (.03)****	.68 (.03)****
Closeness		−.03 (.02)	−.03 (.02)	−.03 (.02)	−.03 (.02)
Social regulation		.02 (.02)	.02 (.02)	.02 (.02)	.02 (.02)
Stress		−.00 (.01)	−.00 (.01)	−.00 (.01)	−.00 (.01)
Closeness × Model		.02 (.10)	.04 (.10)	.02 (.10)	.00 (.10)
Regulation × Model		.03 (.10)	.04 (.10)	.05 (.10)	.05 (.10)
Stress × Model		.01 (.04)	.01 (.04)	.01 (.04)	.01 (.04)
Family × Peer Contexts					
Family Closeness × Peer Model			−.07 (.05)		
Family Regulation × Peer Model			−.14 (.05)**		
Family Stress × Peer Model			.14 (.03)****		
Family Model × Peer Model			.22 (.04)****		
Family × School Contexts					
Family Closeness × School Model				−.02 (.05)	
Family Regulation × School Model				−.08 (.04)*	
Family Stress × School Model				.11 (.02)****	
Family Model × School Model				.19 (.03)****	
Peer × School Contexts					
Peer Closeness × School Model					−.01 (.02)
Peer Regulation × School Model					.07 (.03)†
Peer Stress × School Model					.30 (.21)
Peer Model × School Model					.85 (.32)*
F[a]	64.62****	64.24****	27.73****	19.99****	2.97*

Note. Adolescents are nested in neighborhoods. All hierarchical growth models are computed on five imputed data sets and control for gender, race/ethnicity, family structure, parent education, and high school enrollment.

[a] The F statistic tests the significance of the variables added in each model to the previous model. The microsystem model is compared with a model including only demographic variables.

$^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$. $^{****}p < .0001$.

alcohol misuse. In addition, both family stress and family alcohol use significantly magnified effects of both friends' and schoolmates' alcohol misuse. In the Peers × School model, peer regulation ($p < .10$) and peer alcohol misuse enhanced effects of schoolmates' modeling of alcohol use.

Discussion

Consistent with Bronfenbrenner's ecological perspective on the importance of the multiple social contexts in which lives are embedded, attributes of family, peer, school, and neighborhood contexts uniquely predicted development of adolescent alcohol misuse from age 11 through age 17. The presence of multiple within-context and between-contexts interactions involving constructs from social learning and social control theories largely affirmed Bronfenbrenner's prediction that interactions rather than main effects will be the principal effects demonstrated in ecological research. In addition, our findings partially confirmed our specific expectation that the nature of moderation involving closeness and social regulation would vary depending on whether peers or adults were the referent, whereas stress would always exacerbate alcohol modeling effects. Overall, our findings suggest the appropriateness of an ecological approach for examining adolescent alcohol misuse and support the need, as noted elsewhere (Cook, 2003), to draw on theory to specify the contextual attributes and joint effects examined.

We consider first the within-context findings from the family, peer, school, and neighborhood contexts in the microsystem and exosystem models. In the family and peer contexts, relationships between modeling of alcohol use by others and adolescent alcohol misuse were always moderated by the nature of social bonds. Thus, the social learning effects of modeling could not be estimated in absence of consideration of the quality of relationships. Similarly, effects of relationship attributes identified by social control theory could not be described without reference to alcohol modeling. In addition to this predominance of interactions over main effects, the nature of interactions involving family characteristics was as hypothesized. Parental supervision and adolescent–parent closeness (marginal effect) buffered alcohol modeling effects in the family, whereas family stress exacerbated family alcohol modeling effects. Several previous studies of interactions between family characteristics and family alcohol use found support for mitigating effects of a positive family environment on parent alcohol use (M. P. Farrell et al., 1995; Urberg et al., 2005; Zhang, Welte, & Wieczorek, 1999). Although most prior studies focused on parent–child closeness, our findings extend the buffering effects of parenting to the supervisory domain. Parental supervision and monitoring typically have been examined only as main effects on adolescent alcohol and other substance use.

Two interactions between alcohol modeling and the social control variables characterized the peer context: Peer Alcohol Modeling × Peer Social Regulation and Peer Alcohol Modeling × Peer Stress. In an odd finding, modeling of alcohol misuse by friends was negatively related to alcohol misuse. The interactions were such that at higher levels of peer alcohol use, peer social regulation and peer stress lessened the negative modeling effect. The interactions can be interpreted as consistent with the hypothesized direction of peer effects, in that the risk for alcohol use was increased for youth exposed to higher levels of peer alcohol use and higher social regulation or higher stress compared to those exposed to higher levels of peer alcohol use and lower social regulation or lower stress. Even so, the overall negative effect of peer alcohol use (not present in bivariate analyses), although small, was opposite all expectations and evidently due to the presence of other variables.

In contrast to the family and peer findings, in the school and neighborhood contexts, no significant joint effects of social learning and social bond variables were found. Instead, alcohol modeling by others, whether schoolmates or other adolescents in the neighborhood, was the only significant predictor. It is noteworthy that these modeling effects were detected even after adjusting for friends' modeling of alcohol use. These findings extend the consistently reported association between adolescent and friends' alcohol use to other peer contexts. An implication is that although immediate friendships are central to adolescent alcohol use, the larger and more inclusive school-wide and neighborhood peer environments also clearly matter.

In the mesosystem models, most of the between-context interactions involving characteristics of family and peers, family and school, and peers and school were significant. The expected reverse relationships, depending on whether the family or peers were considered, between social regulation and modeling of alcohol misuse were found. That is, family social regulation buffered effects on adolescent alcohol misuse of alcohol modeling by the adolescent's friends and schoolmates, whereas peer social regulation amplified the modeling effects of alcohol misuse by schoolmates ($p < .10$). The family findings are consistent with studies suggesting that a positive family environment, as indicated by both closeness and supervision, can mitigate negative peer effects (Brook et al., 1986; Brook et al., 1990; Marshal & Chassin, 2000). In addition, and not typically the focus of prior studies, our findings show that the family can amplify negative peer modeling both through family stress (conflict) and through family alcohol use.

Although prior studies of family moderating effects are relatively few, several studies have examined mediation of family alcohol socialization factors. In particular, weak family bonds, as indicated by measures of closeness and supervision, have been shown to lead to association with substance-using peers and then to adolescent use (e.g., Kandel & Andrews, 1987; Simons-Morton & Chen, 2005). Given our conditional effects and those prior mediational findings, perhaps models of moderated mediation should be investigated (Muller, Judd, & Yzerbyt, 2005). That is, investigation could examine whether indirect effects on alcohol misuse of parenting characteristics through peer factors are moderated by attributes of other social contexts. The possibility of moderated mediation indicates the complexity of social contextual relationships and the need for research testing expanded models of relationships among contextual factors. An implication of our findings is that consideration of contingencies in contextual relationship should not be discounted in favor of examining only main effects or mediated relationships.

Taken together, our findings indicate that attributes of all social contexts are relevant to development of adolescent alcohol misuse. Although all contexts were implicated in adolescent alcohol use, the family context emerged as perhaps most important in that all characteristics were associated with adolescent alcohol misuse. As well, all between-context interactions involving the family predicted adolescent alcohol misuse. In contrast to the perception that family influence wanes over adolescence, our findings suggest the enduring influence of the family throughout the adolescent age span examined.

Our findings also indicate the relevance of both social learning and social control theories to adolescent alcohol misuse. Although interrelations between social learning and social control variables generally played out as expected, it is surprising that only the social learning theory variable of alcohol modeling was a significant factor in the school and neighborhood contexts. Perhaps in these more distal, less intimate environments compared to family and friendship groups, adolescents attend to what is most obviously seen—in this case, the alcohol use behavior of others. In addition, or alternatively, perhaps the level of alcohol misuse in this sample was not sufficiently serious to be sensitive to school or neighborhood social controls.

Our analyses are intended to be a specific empirical assessment of Bronfenbrenner's far-reaching perspective on the ecology of human development. The strength of our analyses rests to a great extent in our use of contextual measures that for the most part were constructed independent of our adolescent respondents. Only the measures of the family context and measure of the relationship closure in the peer context were based on adolescent reports. Use of social network analysis allowed us to operationalize peer and school context relational measures with a method specifically intended for measuring relationships. Moreover, due to the data requirements of social network analysis, the measures of the alcohol misuse of friends, schoolmates, and neighbors were based on aggregated self-reports rather than adolescent perceptions. Use of self-reports avoided the bias due to projection effects when adolescent perceptions are used to measure the alcohol use behavior of others (Bauman & Ennett, 1996). At the neighborhood level, our use of parent data and linkage of respondents to census block groups provided measures of social processes in neighborhoods based on multiple informants. The strength of the measures provides reassurance that where expected relationships were not found, the cause is not likely due to inadequate measurement.

Other strengths of our study in the match of methods to Bronfenbrenner's conceptualization of development in context are the application of multilevel modeling to accommodate nested data, examination of developmental trajectories, and use of time-varying measures of the social context variables.

Our study has methodological limitations. Although our analysis of time-varying measures demonstrated that the contribution of social context characteristics to alcohol misuse averaged across all ages examined, we did not test differences at each age in the relationships between the social context variables and the alcohol misuse. Another shortcoming is that our statistical models, although based on longitudinal data, did not allow us to assess temporality of relationships. The models assessed the contemporaneous relationships between the time-varying social context measures and alcohol misuse at each time point assessed (modeled as age); the models did not assess whether the social context attributes at earlier ages predicted alcohol misuse at subsequent ages after controlling for prior involvement. Other statistical models (e.g., autoregressive latent trajectory models) are needed to conduct these analyses and are an essential consideration for future research.

The limitation concerning temporality is most costly in the peer context where we cannot determine whether the relationship between adolescent alcohol misuse and friends' drinking is due to the adolescent's selection of friends or to socialization by those friends. Prior research suggests that both selection and socialization are likely at work (Bauman & Ennett, 1996); indeed, the prediction from Bronfenbrenner's perspective would be that reciprocal processes exist between adolescents and peers. The question is made more complicated by the unexpected negative relationship found in our multivariable models between adolescent alcohol use and friends' alcohol use, although we note that the bivariate relationship was positive. Determining directionality in the family context is also problematic in that adolescent drinking could shape parenting practices or be a cause of family conflict (Ennett, Bauman, Foshee, Pemberton, & Hicks, 2001). Moreover, adolescents could cause their parents' drinking. At the school and neighborhood levels, directionality may flow less ambiguously from the context to the adolescent, although selection by parents into schools and neighborhoods cannot be discounted. With that caveat, our findings contribute to the literature on peer influences on adolescent drinking because exposure to peers' alcohol use at school and in the neighborhood predicted adolescent alcohol misuse net of demographic and family factors, close friends' alcohol misuse, and all other contextual variables.

Besides methodological limitations, our application of Bronfenbrenner's ecology of human development perspective is incomplete in several ways. Perhaps most important is that we did not examine reciprocity between adolescents and their social contexts. As well, we did not assess the influence on adolescent development of the macrosystem, or cultural factors. It is possible that the relationships obtained varied by the race/ethnicity or gender of the respondents or would vary compared to youth in other countries in cross-cultural comparisons. These are areas for future research.

Despite limitations, our study addresses challenges to contextual research in our use of theory to identify specific attributes of contexts and contextual interdependencies for examination. Our findings affirm Bronfenbrenner's social ecological perspective in demonstrating that adolescent alcohol misuse is a socially conditioned behavior. Our findings also affirm the joint relevance of social learning and social control theories in that effects of alcohol use by others were often, although not always, conditional on the nature of social bond. Taken together, our findings support the generality that family, peer, school, and neighborhood social contexts and interdependencies within and between social contexts are significantly implicated in adolescent alcohol misuse.

References

Akers, R. L., Krohn, M. D., Lanza-Kaduce, L., & Radosevich, M. (1979). Social learning and deviant behavior: A specific test of a general theory. *American Sociological Review, 44*, 636–655.

Allison, P. D. (2002). *Missing data*. Thousand Oaks, CA: Sage.

Allison, P. D. (2005). Imputation of categorical variables with PROC MI. In G. S. Nelson (Ed.), *30th Annual SAS Users Group International Conference (SUGI 30)*. Paper 113–30. Retrieved September 2007 from http://www2.sas.com/proceedings/sugi30/toc/html

Andrews, J. A., Hops, H., & Duncan, S. C. (1997). Adolescent modeling of parent substance use: The moderating effect of the relationship with the parent. *Journal of Family Psychology, 11*, 259–270.

Ary, D. V., Tildesley, E., Hops, H., & Andrews, J. A. (1993). The influence of parent, sibling, and peer modeling and attitudes on adolescent use of alcohol. *International Journal of the Addictions*, *28*, 853–880.

Aseltine, R. H., Jr., & Gore, S. L. (2000). The variable effects of stress on alcohol use from adolescence to early adulthood. *Substance Use and Misuse, 35*, 643–668.

Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice-Hall.

Battistich, V., & Hom, A. (1997). The relationship between students' sense of their school as a community and their involvement in problem behaviors. *American Journal of Public Health, 87*, 1997–2001.

Bauman, K. E., & Ennett, S. T. (1996). On the importance of peer influence for adolescent drug use: Commonly neglected considerations. *Addiction, 91*, 185–198.

Bearman, P. S., & Moody, J. (2004). Suicide and friendships among American adolescents. *American Journal of Public Health, 94*, 89–95.

Bloom, B. L. (1985). A factor analysis of self-report measures of family functioning. *Family Process*, *24*, 225–239.

Bogenschneider, K., Wu, M. Y., Raffaelli, M., & Tsay, J. C. (1998). Parent influences on adolescent peer orientation and substance use: The interface of parenting practices and values. *Child Development*, *69*, 1672–1688.

Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist*, *7*, 513–531.

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Cambridge, MA: Harvard University Press.

Brook, J. S., Brook, D. W., Gordon, A. S., Whiteman, M., & Cohen, P. (1990). The psychosocial etiology of adolescent drug use: A family interactional approach. *Genetic, Social, and General Psychology Monographs*, *116*, 111–267.

Brook, J. S., Nomura, C., & Cohen, P. (1989). A network of influences on adolescent drug involvement: Neighborhood, school, peer, and family. *Genetic, Social, and General Psychology Monographs, 115*, 125–145.

Brook, J. S., Whiteman, M., Gordon, A. S., Nomura, C., & Brook, D. W. (1986). Onset of adolescent drinking: A longitudinal study of intrapersonal and interpersonal antecedents. *Advances in Alcohol and Substance Abuse, 5*, 91–110.

Coleman, J. (1988). Social capital in the creation of human capital. *American Journal of Sociology, 94*, 95–120.

Cook, T. D. (2003). The case for studying multiple contexts simultaneously. *Addiction, 98*(Suppl. 1), 151–155.

Curran, P. J., Bauer, D. J., & Willoughby, M. T. (2004). Testing main effects and interactions in latent curve analysis. *Psychological Methods, 9*, 220–237.

Elliott, D. S., Wilson, W. J., Huizinga, D., Sampson, R. J., Elliott, A., & Rankin, B. (1996). The effects of neighborhood disadvantage on adolescent development. *Journal of Research in Crime and Delinquency, 33*, 389–426.

Ennett, S. T., Bauman, K. E., Foshee, V. A., Pemberton, M., & Hicks, K. A. (2001). Parent-child communication about adolescent tobacco and alcohol use: What do parents say and does it affect youth behavior? *Journal of Marriage and the Family, 63*, 48–62.

Ennett, S. T., Bauman, K. E., Hussong, A., Faris, R., Foshee, V. A., DuRant, R. H., et al. (2006). The peer context of adolescent substance use: Findings from social network analysis. *Journal of Research on Adolescence, 16*, 159–186.

Farrell, A. D., King, E. M., White, K. S., & Valois, R. F. (2000). The structure of self-reported aggression, drug use, and delinquent behaviors during early adolescence. *Journal of Clinical Child Psychology, 29*, 282–292.

Farrell, A. D., & White, K. S. (1998). Peer influences and drug use among urban adolescents: Family structure and parent-adolescent relationship as protective factors. *Journal of Consulting and Clinical Psychology, 66*, 248–258.

Farrell, M. P., Barnes, G. M., & Banerjee, S. (1995). Family cohesion as a buffer against the effects of problem-drinking fathers on psychological distress, deviant behavior, and heavy drinking in adolescence. *Journal of Health and Social Behavior, 36*, 377–385.

Giordano, P. C., Cernkovich, S. A., & Pugh, M. D. (1986). Friendships and delinquency. *American Journal of Sociology, 91*, 1170–1202.

Haynie, D. L. (2001). Delinquent peers revisited: Does network structure matter? *American Journal of Sociology, 106*, 1013–1057.

Hirschi, T. (1969). *Causes of delinquency*. Berkeley: University of California Press.

Hops, H., Duncan, T. E., Duncan, S. C., & Stoolmiller, M. (1996). Parent substance use as a predictor of adolescent use: A six-year lagged analysis. *Annals of Behavioral Medicine, 18*, 157–164.

Jackson, C., Henriksen, L., & Foshee, V. A. (1998). The authoritative parenting index: Predicting health risk behaviors among children and adolescents. *Health Education and Behavior, 25*, 319–337.

Kandel, D. B., & Andrews, K. (1987). Processes of adolescent socialization by parents and peers. *International Journal of the Addictions, 22*, 319–342.

Marshal, M. P., & Chassin, L. (2000). Peer influence on adolescent alcohol use: The moderating role of parental support and discipline. *Applied Developmental Science, 4*, 80–88.

Mason, C. A., Cauce, M. A., Gonzales, N., & Hiraga, Y. (1994). Adolescent problem behavior: The effect of peers and the moderating role of father absence and the mother-child relationship. *American Journal of Community Psychology, 22*, 723–743.

Mehta, P., & West, S. G. (2000). Putting the individual back into individual growth curves. *Psychological Methods, 5*, 23–43.

Miyazaki, Y., & Raudenbush, S. W. (2000). Tests for linkage of multiple cohorts in an accelerated longitudinal design. *Psychological Methods, 5*, 44–63.

Moody, J. (2000). *SPAN: SAS programs for analyzing networks* [Web Page]. Retrieved March 2006 from http://www.sociology.ohio-state.edu/jwm/soc_net_methods.htm

Moody, J. (2001). Peer influence groups: Identifying dense clusters in large networks. *Social Networks, 23*, 261–283.

Muller, D., Judd, C. M., & Yzerbyt, V. Y. (2005). When moderation is mediated and mediation is moderated. *Journal of Personality and Social Psychology, 89*, 852–863.

Muthén, L. K., & Muthén, B. O. (1998–2006). *Mplus version 4.21*. Los Angeles: Author.

Parker, E. A., Lichenstein, R. L., Schulz, A. J., Israel, B. A., Schork, M. A., Steinman, K. J., et al. (2001). Disentangling measures of individual perceptions of community social dynamics: Results of a community survey. *Health Education & Behavior, 28*, 462–486.

Perry, C. L., Kelder, S. H., & Komro, K. (1993). A world of adolescents: Family, peers, schools, and the community. In S. G. Millstein, A. C. Peterson, & E. O. Nightingale (Eds.), *Promoting the health of adolescents: New directions for the twenty-first century* (pp. 73–96). New York: Oxford University Press.

Petraitis, J., Flay, B. R., & Miller, T. Q. (1995). Reviewing theories of adolescent substance use: Organizing pieces in the puzzle. *Psychological Bulletin, 117*, 67–86.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology, 25*, 111–163.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). London: Sage.

Richter, L., & Richter, D. M. (2001). Exposure to parental tobacco and alcohol use: Effects on children's health and development. *American Journal of Orthopsychiatry, 71*, 182–203.

Ross, C. E., & Jong, S. J. (2000). Neighborhood disorder, fear, and mistrust: The buffering role of social ties with neighbors. *American Journal of Community Psychology, 28*, 401–420.

Sampson, R. J., Raudenbush, S. W., & Earls, F. (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science, 277*, 918–924.

SAS. (2002–2003). *SAS 9.1. for Windows*. Cary, NC: SAS Institute.

Schaefer, J. L. (2001). Multiple imputation with PAN. In L. M. Collins & A. Sayer (Eds.), *New methods for the analysis of change*. Washington, DC: American Psychological Association.

Simons-Morton, B., & Chen, R. (2005). Latent growth curve analyses of parent influences on drinking progression among early adolescents. *Journal of Studies on Alcohol, 66*, 5–13.

Thissen, D., Chen, W. H., & Bock, R. D. (2003). *Multilog* (version 7). Lincolnwood, IL: Scientific Software International.

Thissen, D., Nelson, L., Rosa, K., & McLeod, L. D. (2001). Item response theory for items scored in two categories. In D. Thissen & H. Wainer (Eds.), *Test scoring* (pp. 141–186). Mahwah, NJ: Erlbaum.

Thissen, D., & Orlando, M. (2001). Item response theory for items scored in two categories. In D. Thissen & H. Wainer (Eds.), *Test scoring* (pp. 73–140). Mahwah, NJ: Erlbaum.

Urberg, K., Goldstein, M. S., & Toro, P. A. (2005). Supportive relationships as a moderator of the effects of parent and peer drinking on adolescent drinking. *Journal of Research on Adolescence, 15*, 1–19.

Webb, J. A., & Baer, P. E. (1995). Influence of family disharmony and parental alcohol use on adolescent social skills, self-efficacy, and alcohol use. *Addictive Behaviors, 20*, 127–135.

Zhang, L., Welte, J. W., & Wieczorek, W. F. (1997). Peer and parental influences on male adolescent drinking. *Substance Use and Misuse, 32*, 2121–2136.

Zhang, L., Welte, J. W., & Wieczorek, W. F. (1999). The influence of parental drinking and closeness on adolescent drinking. *Journal of Studies on Alcohol, 60*, 245–251.

Child Development, November/December 2008, Volume 79, Number 6, Pages 1792 – 1801

Trumping Shame by Blasts of Noise: Narcissism, Self-Esteem, Shame, and Aggression in Young Adolescents

Sander Thomaes
VU University Amsterdam and Utrecht University

Brad J. Bushman
University of Michigan and VU University Amsterdam

Hedy Stegge and Tjeert Olthof
VU University Amsterdam

This experiment tested how self-views influence shame-induced aggression. One hundred and sixty-three young adolescents (M = 12.2 years) completed measures of narcissism and self-esteem. They lost to an ostensible opponent on a competitive task. In the shame condition, they were told that their opponent was bad, and they saw their own name at the bottom of a ranking list. In the control condition, they were told nothing about their opponent and did not see a ranking list. Next, participants could blast their opponent with noise (aggression measure). As expected, narcissistic children were more aggressive than others, but only after they had been shamed. Low self-esteem did not lead to aggression. In fact, narcissism in combination with high self-esteem led to exceptionally high aggression.

Violent aggressive behavior in youth leaves scars on perpetrators, victims, and society at large. Among the many factors that contribute to youth aggression, the self-regard of perpetrators has been a theoretically important but empirically controversial cause. For many years, the prevailing view has held that aggressive youth have low self-esteem (e.g., California Task Force to Promote Self-Esteem and Personal and Social Responsibility, 1990; Carr, 1999; Heide, 1997; Kaplan, 1975; Keith, 1984; Willock, 1987). Applied and practical efforts have also focused on low self-esteem as a cause of violence. For example, following a series of incidents in which schoolchildren fired guns and killed their classmates at various American schools, several organizations (including the U.S. Department of Education) prepared lists of alleged warning signals to be used to identify youth who might be relatively likely to engage in such destructive violence, and nearly all the lists included low self-esteem as a significant risk factor (e.g., Lord, 1999).

Despite this apparent consensus, no compelling theoretical rationale existed to explain why low self-esteem would cause aggression. Even more problematic, a persuasive body of empirical evidence was lacking. Although there have been a few exceptions (Lochman & Dodge, 1994), most studies involving elementary school-aged children found no concurrent or prospective links between low self-esteem and aggression (Gresham, MacMillan, Bocian, Ward, & Forness, 1998; Hymel, Rubin, Rowden, & LeMare, 1990; Kupersmidt & Patterson, 1991). In fact, several studies found that aggressive children have inflated rather than deflated self-views (Brendgen, Vitaro, Turgeon, Poulin, & Wanner, 2004; David & Kistner, 2000; Zakriski & Coie, 1996) and that high self-esteem is a risk factor for exacerbating aggression problems over time (Menon et al., 2007).

One possible explanation of these findings is that low self-esteem may only cause aggression in adolescents because they are more concerned than younger children about maintaining desired self-images (Harter, 2006; Rosenberg, 1986). Contrary to that reasoning, however, the bulk of studies involving adolescents also found no link between low self-esteem and aggression (East & Rook, 1992; Esposito, Kobak, & Little, 2005; Olweus, 1994; Prinstein, Boergers, & Vernberg, 2001; Salmivalli, Kaukiainen, Kaistaniemi, & Lagerspetz, 1999; but see Donnellan, Trzesniewski, Robins, Moffitt, & Caspi, 2005, for an exception). As it

---

This research was supported by a Fulbright scholarship to S.T. We are grateful to the children and staff of the Brighton School District in Michigan for their kind cooperation. We also thank Marcel van Aken, Roy Baumeister, Keith Campbell, and Bram Orobio de Castro for their valuable comments on an earlier version of this article.

Correspondence concerning this article should be addressed to Sander Thomaes, Department of Developmental Psychology, Utrecht University, P.O. Box 80.140, 3584 CS, Utrecht, The Netherlands. Electronic mail may be sent to s.thomaes@uu.nl.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7906-0018

stands, there is little reason to adhere to the view that low self-esteem causes aggressive behavior in youth.

Narcissism and Aggression

In a comprehensive literature review, Baumeister, Smart, and Boden (1996) rejected the view that low self-esteem causes aggression. They proposed instead that violence most commonly occurs when inflated views of self and unstable beliefs in personal superiority are threatened. These conceptions of excessive self-love are relevant to narcissism, a term that comes from the Greek myth about a handsome young man who falls in love with his own reflection in the water. In its extreme form, narcissism is a personality disorder that involves grandiose views of self, an inflated sense of entitlement, and exploitative attitudes toward others (American Psychiatric Association, 1994). Most current psychological research focuses on "normal narcissism," operationalized as a trait on which people in the general population vary (Morf & Rhodewalt, 2001; Raskin & Terry, 1988). Intuition suggests that narcissism is equivalent to excessively high self-esteem, but research showed that narcissism and self-esteem are not strongly correlated (Brown & Zeigler-Hill, 2004). Narcissists are self-absorbed and arrogant, but they differ in feelings of self-assurance and sufficiency (reflecting higher levels of self-esteem) or self-doubt and insufficiency (reflecting lower levels of self-esteem; Cain, Pincus, & Ansell, 2007; Dickinson & Pincus, 2003; Rose, 2002; Wink, 1991).

The link between narcissism and aggression has been firmly established in adults (e.g., Bushman & Baumeister, 1998; Bushman, Bonacci, van Dijk, & Baumeister, 2003; Donnellan et al., 2005; Stucke & Sporer, 2002; Twenge & Campbell, 2003). Preliminary evidence suggests that narcissists with high self-esteem are particularly prone to behave aggressively (Papps & O'Carroll, 1998; Wink, 1991). Unfortunately, few studies have examined the effects of narcissism and self-esteem on aggression in youth. This lack of emphasis is surprising because childhood is the time when the foundation for lifelong aggressive or nonaggressive behavior styles is laid (e.g., Loeber & Hay, 1997). For both clinical and theoretical reasons, it is important to study the psychological mechanisms underlying aggressive inclinations at a time that they unfold. One cause of the paucity of research on narcissistic aggression in youth has been the absence, until recently, of a scale designed to measure early manifestations of narcissism. Recent developments indicate, however, that the construct of narcissism can be reliably identified and distinguished from related personality dimensions in older children and adolescents (C. T. Barry, Frick, & Killian, 2003; Frick, Bodin, & Barry, 2000). We recently developed the Childhood Narcissism Scale, a reliable and valid measure of narcissistic traits in normal child and adolescent populations (Thomaes, Stegge, Bushman, Olthof, & Denissen, 2008).

Shame-Induced Aggression

Violence and aggression often occur when an individual's pride, reputation, or self-esteem has been impugned or threatened. In late childhood and adolescence, such situations are typically experienced as shameful (Nishina & Juvonen, 2005; Olthof, Ferguson, Bloemers, & Deij, 2004; Reimer, 1996). Shameful events often involve the public exposure of some failure or shortcoming (Olthof, Schouten, Kuiper, Stegge, & Jennekens-Schinkel, 2000; Smith, Webster, Parrott, & Eyre, 2002). When shamed, people are painfully aware that others might think they are flawed (Lewis, 1971; Tangney & Dearing, 2002). Importantly, this awareness of others' disapproval is easily internalized as a global condemnation of the self (e.g., "I am a bad and worthless person"). Over the course of late childhood, such self-condemning negative self-appraisals become a more pronounced part of the shame experience (Ferguson, Stegge, & Damhuis, 1991). Gradually, shameful events come to constitute a more serious threat to self-esteem.

How do people behave when they are shamed? Shameful events may cause people to withdraw and hide from social contact (e.g., Lindsay-Hartz, De Rivera, & Mascolo, 1995). Alternatively, shameful events may cause people to lash out aggressively against others. Across the life span, shame-prone individuals are predisposed to externalize blame, to feel anger, and to behave aggressively (Tangney, Wagner, Hill-Barlow, Marschall, & Gramzow, 1996). Situationally induced shame produces similar reactions (Bennett, Sullivan, & Lewis, 2005; Thomaes, Stegge, & Olthof, 2005). Shame-induced aggression may serve an ego-protective function (Tangney & Dearing, 2002). By directing blame and anger on others, people can prevent their self-esteem from (further) damage. Aggression shifts attention away from painful awareness of a devalued self. Also, by asserting the dominant aggressive stance, people can reaffirm the self and "save face" in front of others. In summary, shame-induced aggression may originate from the basic human motive to preserve self-esteem.

Self-Views and Shame-Induced Aggression

If the traditional view that low self-esteem causes aggression is true, one would predict that youth with low self-esteem should be particularly aggressive when shamed because shameful events make them feel even more inferior. This prediction, however, is inconsistent with what we know about the motivations that surround self-esteem. Self-verification theory holds that people generally try to maintain consistent self-appraisals and dislike changing their self-views (Swann & Read, 1981). From this perspective, youth with low self-esteem should be relatively untouched by shameful events because their habitual self-appraisals are already low. In contrast, youth with inflated narcissistic self-views should be more vulnerable than others to shameful events because they are highly motivated to protect their inflated self from being damaged. Indeed, vulnerability to shame has been described as a key component of narcissism (Morrison, 1989; Tracy & Robins, 2004). These notions are consistent with the empirical findings in adult samples showing that narcissists are aggressive following ego threat.

Present Experiment

A between-subjects experimental design was used to examine how self-views influence aggressive responses to induced shame. Participants were 10–13 years old. From several developmental perspectives, this is an ideal age period for the purposes of this experiment. First, in early adolescence, shame is both a frequent and an aversive emotional experience due to developmental increases in self-consciousness (Ryan & Kuczkowski, 1994; Simmons, Rosenberg, & Rosenberg, 1973), the ability to view the self from others' perspective (Harter, 2006), and the ability to make global negative evaluations of the self (Ferguson et al., 1991). Second, early adolescence is a time when children become increasingly concerned about maintaining worth and approval, and self-protective motives come to exert a stronger influence on their behavior (Harter, 2006; Rosenberg, 1986). Third, early adolescence is an ideal time to measure childhood narcissism. Young children typically hold unrealistically positive self-views because they lack the abilities to differentiate their actual self from their ideal self and to base their self-views on social comparisons (Harter, 2006). By age 10, most youth have overgrown age-normative overestimation of competence that likely is a prerequisite for the meaningful assessment of individual differences in childhood narcissism (Thomaes et al., 2008). Fourth, early adolescence is a time when children's self-views are still relatively malleable before they become crystallized in late adolescence and adulthood (Trzesniewski, Donnellan, & Robins, 2003). Thus, early adolescence may be a critical age period to initiate aggression interventions targeted at changing self-views. This experiment tested whether a link between self-views and aggression has already been established by then.

Aggressive behavior was examined in a *situational context* of shame. We did not address how *felt* shame was involved in youth's aggressive behavior. Several authors have noted that although feelings of shame may mediate shame-induced aggression, these feelings are not necessarily consciously experienced (e.g., Lewis, 1971; Robins, Tracy, & Shaver, 2001). For conceptual clarity, we chose to focus on shame-induced aggression on a situational level.

We used a shame manipulation based on the easy task failure paradigm (e.g., Lewis, Alessandri, & Sullivan, 1992). People who fail an easy task are especially likely to experience shame. Participants in our experiment failed a competitive reaction time task. They were randomly assigned to shame or no-shame control conditions. Participants in the shame condition were told that their opponent was one of the slowest contestants tested so far, and they saw their own name at the bottom of a ranking list posted on the bogus FastKid! Web page (below their opponent's name). The Internet rankings highlighted public exposure, which should enhance feelings of shame (Smith et al., 2002). Participants in the control condition were told nothing about their opponent's abilities and did not see the bogus Web page. A validation study showed this shame manipulation to be highly effective (Thomaes et al., 2005). The shame condition elicits feelings of shame in young adolescents, whereas the control condition does not. Next, participants were given a chance to blast their opponent with loud noise through headphones (the aggression measure). We predicted that narcissistic youth would behave most aggressively, but only in the shame condition. We did not predict high levels of aggression in youth with low self-esteem. On the contrary, on the basis of prior research, we predicted that high self-esteem would enhance narcissistic aggression (Papps & O'Carroll, 1998; Wink, 1991).

Method

Participants

Participants were 163 young adolescents (54% boys) from two public middle schools serving middle- and upper-middle-class neighborhoods in southeastern Michigan. In the school district, 8% of students are eligible for reduced-price lunch programs. Participants ranged in age from 10 to 13 years ($M = 12.2$, $SD = 0.6$). Almost all were Caucasians (96%). To participate, adolescents received informed parental consent (29% of parents consented) and gave their own assent (98% of adolescents assented). Participants received a small gift (e.g., mechanical pens, markers) in exchange for their voluntary participation.

Self-View Questionnaire

A few weeks prior to the experiment ($M = 3$ weeks, range = 1–5 weeks), participants completed self-report measures of narcissism and self-esteem at their school. Narcissism was measured using the reliable and valid Childhood Narcissism Scale (Thomaes et al., 2008). This 10-item scale assesses grandiose views of self, inflated feelings of superiority and entitlement, and exploitative interpersonal attitudes. Sample items include "Without me, our class would be much less fun," "Kids like me deserve something extra," and "I often succeed in getting admiration." Items are rated along a 4-point scale ranging from 0 (*not at all true*) to 3 (*completely true*). Responses were summed, with higher scores indicating higher levels of narcissism. In the present experiment, the alpha coefficient for the scale was .76.

Self-esteem was measured using the 6-item *global self-worth* subscale of the Self-Perception Profile for Children (Harter, 1985). This scale assesses the extent to which participants are satisfied with themselves and the way they are leading their lives. Sample items include "Some kids like the kind of person they are" and "Some kids are not very happy with the way they do a lot of things." Following others (e.g., Brendgen et al., 2004), we used a 4-point scale response format ranging from 0 (*I am not like these kids at all*) to 3 (*I am exactly like these kids*). After reverse scoring negatively worded items, responses were summed, with higher scores indicating higher levels of self-esteem. The alpha coefficient for the scale was .72.

Procedure

Participants were tested individually in a quiet room at their school. They were told that they would be competing on an Internet reaction time game called FastKid! against an opponent of the same sex and age from a school in Columbus, OH. In reality, there was no opponent and the computer controlled all events. Participants were told that FastKid! consisted of two 5-trial rounds, each with a bonus. The first-round bonus was the ability to send a written message to the opponent. The second-round bonus was the ability to blast the opponent with loud white noise (sounds like radio static) through headphones after winning a trial. Through a rigged lottery, the opponent had the bonus in the first round, whereas the participant had the bonus in the second round. Participants were given samples of noise they could set for their opponent. The noise levels ranged from 55 (Level 1) to 100 dB (Level 10) in 5-dB increments. The maximum noise level, 100 dB, is about the same intensity as a smoke alarm. A nonaggressive no-noise setting (Level 0) was also included.

Participants were randomly assigned to shame or no-shame control conditions. In the shame condition, participants were told that they were lucky to compete against one of the worst players thus far. The experimenter then logged onto the fictitious FastKid! Web site and showed participants their opponent's name at the bottom of the ranking list. The experimenter said, "This means you should win easily!" Participants were told that immediately after the first round of the game, new rankings would appear on the very popular FastKid! Web site and that their own name would be included in those rankings. After competing with the opponent on the first five reaction time trials, a message appeared on screen that said, "Sorry (participant's name), you lost!" The opponent then sent the participant a message that said, "Can't wait to see the rankings!" Then, the new rankings showed the participant's name at the bottom of the list, beneath the opponent's name. The control condition was similar to the shame condition, with three exceptions. First, participants received no information about how good their opponent was (and saw no rankings on the Web site before the game). Second, participants saw no rankings after the game. Third, the opponent's message said, "Huh?! Is the first round finished already?" We used a losing control condition because we wanted to test the effects of shame above and beyond the effects of mere disappointment or frustration from losing a game.

In the second round of the game, participants had the "noise bonus," so they could blast their opponent with loud noise after winning a trial. Prior to each of the five trials of Round 2, participants set the noise level their opponent would receive if the opponent lost. After each trial, participants were informed whether they had won (i.e., Trials 1, 2, 4, and 5) or lost (i.e., Trial 3) that trial. To obtain an aggression measure unconfounded by the (nonmanipulation) effect of losing Trial 3, the average level of noise set for the opponent across the first three trials was used to measure aggression. The alpha coefficient for the aggression measure was .85. Finally, participants were thoroughly debriefed to remove lingering effects of the manipulations and were given a small gift.

Results

Preliminary Analyses

Sex differences and age. Boys were significantly more aggressive than girls, $F(1, 161) = 8.37, p < .01, d = 0.45$. Aggression was not affected by age, $t(161) = -0.33, p > .74, b = -0.07, \beta = -.03$. Because there were also no interactions involving sex or age, the data from boys and girls of different ages were combined for subsequent analyses.

Equivalence of experimental conditions. Narcissism and self-esteem scores did not differ in the shame and no-shame groups ($ps > .30$). Thus, random assignment to conditions was effective. Narcissism and self-esteem were only weakly correlated ($r = .09$).

Primary Analyses

Table 1 contains the descriptive statistics for the experiment. Data were analyzed using hierarchical multiple regression analysis. The dependent variable was aggressive behavior, defined as the average intensity of noise participants gave their ostensible opponent. The main effects for condition, narcissism, and self-esteem were entered in Step 1; the two-way interactions involving these variables were entered in Step 2; and the three-way interaction was entered in Step 3. Narcissism and self-esteem scores were centered to reduce multicollinearity (Aiken & West, 1991; Jaccard & Turrisi, 2003). A maximum variance inflation factor (VIF) greater than 10 indicates that multicollinearity may be unduly influencing the least squares estimates (e.g., Neter, Wasserman, & Kutner, 1990). The maximum VIF in the regression analysis was 3.26, indicating that multicollinearity was not a problem.

Table 1

Means and Standard Deviations in the Shame and No-Shame Control Conditions

		Shame (N = 83)	Control (N = 80)
	Range	M (SD)	M (SD)
Narcissism	2–26	12.04 (4.32)	11.29 (4.92)
Self-esteem	3–18	13.60 (2.89)	13.51 (3.38)
Aggression	1–10	7.06 (2.51)	7.02 (2.66)
Age (months)	131–166	146 (8)	146 (8)

The analysis revealed a main effect for narcissism, $t(159) = 2.01, p < .05, b = 0.41, \beta = .16$. This main effect, however, was qualified by a significant interaction between narcissism and condition, $t(156) = 2.03, p < .05, b = 0.82, \beta = .21$ (see Figure 1). As expected, narcissism was positively related to aggression when participants were shamed, $t(81) = 2.91, p < .01, b = 0.89, \beta = .31$. Narcissism was not related to aggression when participants were not shamed, $t(78) = 0.04, p > .96, b = 0.01, \beta = .01$.

Although self-esteem did not directly influence aggression levels, on the basis of prior research, we anticipated that narcissism in combination with high self-esteem would lead to exceptionally high levels of aggression in the shame condition (Papps & O'Carroll, 1998; Wink, 1991). As expected, there was a significant Narcissism × Self-Esteem × Condition interaction, $t(155) = 2.00, p < .05, b = 0.84, \beta = .19$. To interpret the three-way interaction, we examined the two-way interactions between narcissism and self-esteem separately for the shame and no-shame control conditions. As expected, narcissism and self-esteem interacted to influence aggression in the shame condition, $t(79) = 2.65, p < .01, b = 0.91, \beta = .28$ (see Figure 2; high values of narcissism and self-esteem were 1 SD above the mean and low values were 1 SD below the mean; Aiken & West, 1991). Figure 2 shows that narcissism and aggression were

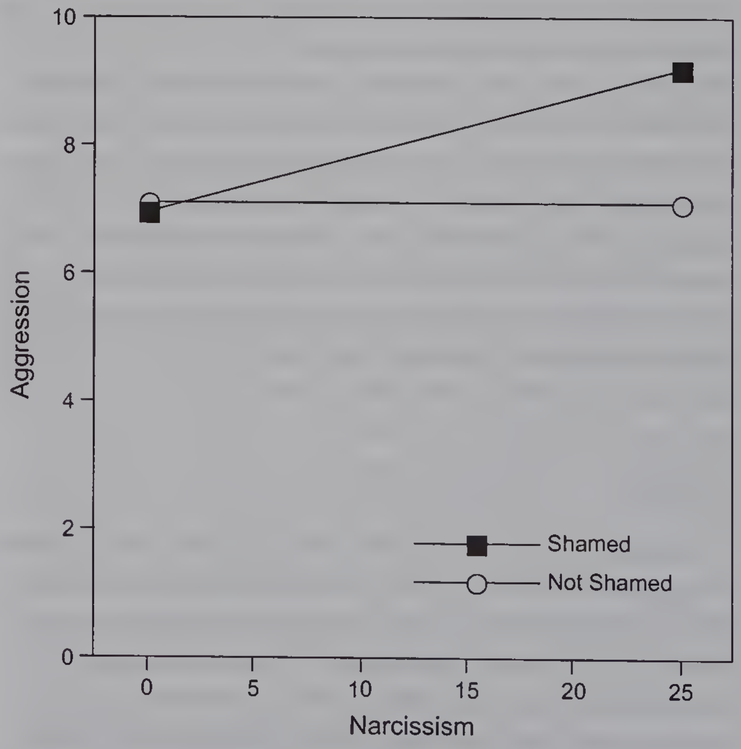


Figure 1. Relationship between narcissism and aggression for participants in the shame and the no-shame control conditions.

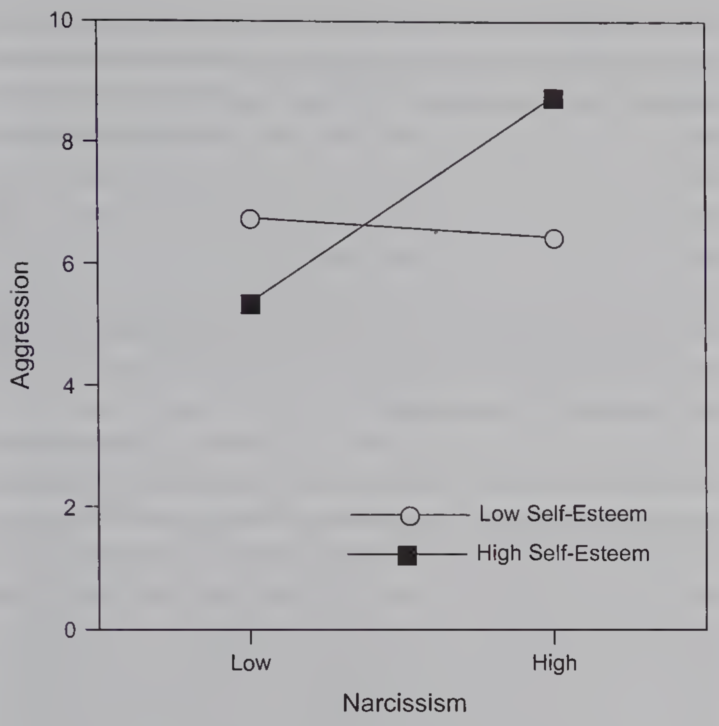


Figure 2. Relationship between narcissism and aggression for participants with high and low self-esteem in the shame condition.
Note. High values of narcissism and self-esteem are 1 *SD* above the mean; low values of narcissism and self-esteem are 1 *SD* below the mean.

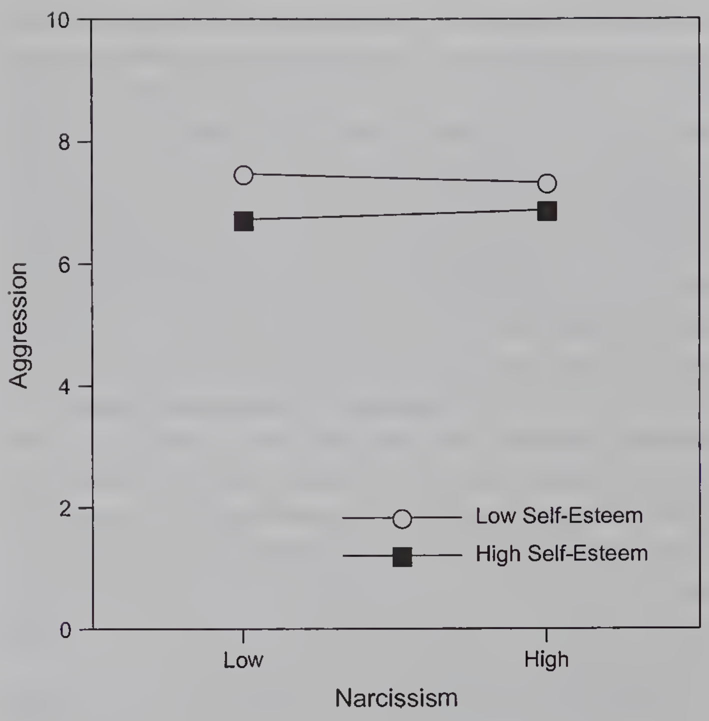


Figure 3. Relationship between narcissism and aggression for participants with high and low self-esteem in the no-shame control condition.
Note. High values of narcissism and self-esteem are 1 *SD* above the mean; low values of narcissism and self-esteem are 1 *SD* below the mean.

strongly associated in shamed youth with high self-esteem, $t(79) = 4.00, p < .001, b = 1.69, \beta = .59$. In contrast, narcissism and aggression were not associated in shamed youth with low self-esteem, $t(79) = -0.29, p > .77, b = -0.14, \beta = -.05$. As expected, narcissism and self-esteem did not interact to influence aggression in the no-shame control condition, $t(76) = 0.30, p > .76, b = 0.07, \beta = .04$. As shown in Figure 3, narcissism and aggression were not associated in nonshamed youth with high or low self-esteem, $t(76) = 0.24, p > .81, b = 0.09, \beta = .04$ and $t(76) = -0.17, p > .86, b = -0.06, \beta = -.03$, respectively.

Discussion

The present experiment examined how narcissism and self-esteem influence young adolescents' shame-induced aggressive behavior. As predicted, narcissistic youth were more aggressive than others, but only after they had been shamed. Narcissists seem highly motivated to create and maintain a grandiose view of self. They tend to interpret social situations in terms of how they reflect on the self, and they engage in self-regulatory strategies to protect self-esteem when they need to (Morf & Rhodewalt, 2001). As shameful situations constitute a threat to grandiosity, narcissistic shame-induced aggression can likely be viewed as defensive effort to maintain self-worth.

No support was found for the traditional view that low self-esteem underlies aggression. In fact, that view was contradicted by the finding that high self-esteem (not low self-esteem) increased narcissistic shame-induced aggression. One explanation for this finding, based on self-verification theory, is that narcissistic youth with high self-esteem are more vulnerable to shameful events than are youth with low self-esteem. Another explanation is that narcissistic youth with high and low self-esteem do not differ in their vulnerability to shameful events, but they do differ in the way they deal with those events. This latter explanation is consistent with research showing that narcissists' interpersonal orientation tends to vary along with their level of self-esteem (e.g., Cain et al., 2007; Wink, 1991). *Overt narcissists*, who have high self-esteem, have been described as extraverts marked by a dominant and aggressive interpersonal orientation. *Covert narcissists*, who have much lower self-esteem (i.e., their self-absorption co-occurs with feelings of self-doubt and insufficiency), have been described as "worriers" marked by an anxious and internalizing interpersonal orientation. They may feel aggrieved and hurt when shamed, but in our experiment, they were not aggressive.

Researchers have argued that aggression is an appealing behavioral alternative to shamed individuals because it serves an ego-protective function (Tangney & Dearing, 2002). Aggression provides immediate relief from the pain of shame, which is a tempting benefit in the short run. In the long run, however, predispositions to behave aggressively when shamed may have serious costs. Children who persistently deflect the painful feelings associated with their flaws and shortcomings may become less motivated to overcome them. Consequently, they may become less well adapted to the demands of their social environment. Also, affectively aggressive children are unpopular with peers (e.g., Price & Dodge, 1989; Prinstein & Cillessen, 2003). Thus, aggressive behaviors meant to discard shame in the short run may ironically increase children's liability to be the target of victimization in the long run.

The present experiment contributes to the literature in several ways. To our knowledge, it is the first study to examine the link between youth's self-views and their actual aggressive behavior in an in vivo situational context. Other studies have relied on measures of reported aggression that often include more diffusely defined antisocial acts such as lying or stealing (e.g., Donnellan et al., 2005). Our findings are consistent with those of previous work that found that narcissism is linked to reported aggression and conduct problems in youth (C. T. Barry et al., 2003; T. D. Barry et al., 2007). Also, the present experiment highlights the importance of shame as an emotional context for examining the link between self-views and aggression. Most important, this experiment shows the value of differentiating among different forms of self-view (Salmivalli, 2001). Most previous research involving children has relied exclusively on measures of self-esteem, which by itself is an unreliable predictor of aggression. This study indicates that narcissism is an important predictor of aggression, especially in the context of shame.

Limitations and Future Research

We have argued that early adolescence is an ideal age period for the present experiment. However, our developmental focus limits the ability to make generalizations to other age periods. Future research should test whether the link between self-views and aggression is moderated by age. Such knowledge would help clinicians to identify critical age periods to target the self-views underlying children's aggressive behavior.

The parental consent rate for our experiment was low (i.e., 29%). This may have occurred because the letter from the principal endorsing our study was accidentally not included with the consent form. Low parental consent does not threaten the internal validity of our experiment because participants were randomly assigned to shame or no-shame groups. We cannot, however, exclude the possibility that self-selected biased sampling has influenced the magnitude of the experimental effects observed. Future research may assess the extent to which our experimental findings generalize to other community and noncommunity samples.

Experimental aggression research is sometimes criticized for using laboratory paradigms that are supposedly unrepresentative of "real-world" aggression. The validity of laboratory aggression paradigms (including the noise blast procedure we used) has been supported in two meta-analyses. One meta-analysis demonstrated high levels of convergence across a wide range of laboratory aggression measures (Carlson, Marcus-Newhall, & Miller, 1989). The other meta-analysis showed that "real" and laboratory measures of aggression are influenced in similar ways by situational variables (e.g., provocation) and by individual difference variables (e.g., trait aggressiveness; Anderson & Bushman, 1997).

In contrast to the adult literature, aggression in the context of self-esteem threats such as shame is rarely examined in the child literature. We believe this is unfortunate because such threats are common experiences (particularly in early adolescence) known to elicit aggression in subsets of children (Dodge, Coie, & Lynam, 2006). In this experiment, we manipulated a situation of self-attributable shortcoming. However, shame and other self-esteem threats can also result from shortcomings pinpointed by others. In fact, peer harassment among schoolchildren typically involves damaging others' self-esteem or status (Galen & Underwood, 1997; Nishina & Juvonen, 2005). Continued research on shame-induced aggression is needed to obtain a more complete view on the emotional processes involved in children's and adolescents' aggression.

Conclusion

Many efforts to reduce violence and aggression in youth have relied on boosting self-esteem (e.g., Kusché & Greenberg, 1994; Ringwalt, Graham, Paschall, Flewelling, & Browne, 1996). Although intuitively plausible, there are no clear theoretical or empirical reasons why boosting self-esteem should be effective in reducing aggression. In fact, the present experiment suggests that practices aimed at boosting self-esteem may even bring substantial costs. If these practices cultivate the inflated views of self that are characteristic of narcissism, they are likely to increase (rather than to decrease) the aggressive behavior of youth at risk.

References

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Anderson, C. A., & Bushman, B. J. (1997). External validity of "trivial" experiments: The case of laboratory aggression. *Review of General Psychology, 1*, 19–41.

Barry, C. T., Frick, P. J., & Killian, A. L. (2003). The relation of narcissism and self-esteem to conduct problems in children: A preliminary investigation. *Journal of Clinical Child and Adolescent Psychology, 32*, 139–152.

Barry, T. D., Thompson, A., Barry, C. T., Lochman, J. E., Adler, K., & Hill, K. (2007). The importance of narcissism in predicting proactive and reactive aggression in moderately to highly aggressive children. *Aggressive Behavior, 33*, 185–197.

Baumeister, R. F., Smart, L., & Boden, J. M. (1996). Relation of threatened egotism to violence and aggression: The dark side of high self-esteem. *Psychological Review, 103*, 5–33.

Bennett, D. S., Sullivan, M. W., & Lewis, M. (2005). Young children's adjustment as a function of maltreatment, shame, and anger. *Child Maltreatment, 10*, 311–323.

Brendgen, M., Vitaro, F., Turgeon, L., Poulin, F., & Wanner, B. (2004). Is there a dark side of positive illusions? Overestimation of social competence and subsequent adjustment in aggressive and nonaggressive children. *Journal of Abnormal Child Psychology, 32*, 305–320.

Brown, R. P., & Zeigler-Hill, V. (2004). Narcissism and the non-equivalence of self-esteem measures: A matter of dominance? *Journal of Research in Personality, 38*, 585–592.

Bushman, B. J., & Baumeister, R. F. (1998). Threatened egotism, narcissism, self-esteem, and direct and displaced aggression: Does self-love or self-hate lead to violence? *Journal of Personality and Social Psychology, 75*, 219–229.

Bushman, B. J., Bonacci, A. M., van Dijk, M., & Baumeister, R. F. (2003). Narcissism, sexual refusal, and aggression: Testing a narcissistic reactance model of sexual coercion. *Journal of Personality and Social Psychology, 84*, 1027–1040.

Cain, N. M., Pincus, A. L., & Ansell, E. B. (2007). Narcissism at the crossroads: Phenotypic description of pathological narcissism across clinical theory, social/personality psychology, and psychiatric diagnosis. *Clinical Psychology Review, 28*, 638–656.

California Task Force to Promote Self-Esteem and Personal and Social Responsibility. (1990). *Toward a state of esteem*. Sacramento: California State Department of Education.

Carlson, M., Marcus-Newhall, A., & Miller, N. (1989). Evidence for a general construct of aggression. *Personality and Social Psychology Bulletin, 15*, 377–389.

Carr, A. (1999). *The handbook of child and adolescent clinical psychology: A contextual approach*. London: Routledge.

David, C. F., & Kistner, J. A. (2000). Do positive self-perceptions have a "dark-side"? Examination of the link between perceptual bias and aggression. *Journal of Abnormal Child Psychology, 28*, 327–337.

Dickinson, K. A., & Pincus, A. L. (2003). Interpersonal analysis of grandiose and vulnerable narcissism. *Journal of Personality Disorders, 17*, 188–207.

Dodge, K. A., Coie, J. D., & Lynam, D. (2006). Aggression and antisocial behavior in youth. In W. Damon & R. Lerner (Series Eds.) & N. Eisenberg (Vol. Ed.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (pp. 719–788). New York: Wiley.

Donnellan, M. B., Trzesniewski, K. H., Robins, R. W., Moffitt, T. E., & Caspi, A. (2005). Low self-esteem is related to aggression, antisocial behavior, and delinquency. *Psychological Science, 16*, 328–335.

East, P. L., & Rook, K. S. (1992). Compensatory patterns of support among children's peer relationships: A test using school friends, nonschool friends, and siblings. *Developmental Psychology, 28*, 163–172.

Esposito, A. J., Kobak, R., & Little, M. (2005). Aggression and self-esteem: A diary study of children's reactivity to negative interpersonal events. *Journal of Personality, 73*, 887–906.

Ferguson, T. J., Stegge, H., & Damhuis, I. (1991). Children's understanding of guilt and shame. *Child Development, 62*, 827–839.

Frick, P. J., Bodin, S. D., & Barry, C. T. (2000). Psychopathic traits and conduct problems in community and clinic-referred samples of children: Further development of the psychopathy screening device. *Psychological Assessment, 12*, 382–393.

Galen, B. R., & Underwood, M. K. (1997). A developmental investigation of social aggression among children. *Developmental Psychology, 33*, 589–600.

Gresham, F. M., MacMillan, D. L., Bocian, K. M., Ward, S. L., & Forness, S. R. (1998). Comorbidity of hyperactivity-impulsivity-inattention and conduct problems: Risk factors in social, affective, and academic domains. *Journal of Abnormal Child Psychology, 26*, 293–406.

Harter, S. (1985). *The self-perception profile for children: Revision of the perceived competence scale for children*. Denver, CO: University of Denver.

Harter, S. (2006). The self. In W. Damon & R. Lerner (Series Eds.) & N. Eisenberg (Vol. Ed.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (pp. 505–570). New York: Wiley.

Heide, K. M. (1997). Juvenile homicide in America: How can we stop the killing? *Behavioral Sciences and the Law, 15*, 203–220.

Hymel, S., Rubin, K. H., Rowden, L., & LeMare, L. (1990). Children's peer relationships: Longitudinal prediction of internalizing and externalizing problems from middle to late childhood. *Child Development*, *61*, 2004 – 2021.

Jaccard, J., & Turrisi, R. (2003). *Interaction effects in multiple regression* (2nd ed.). Newbury Park, CA: Sage.

Kaplan, H. B. (1975). *Self-attitudes and deviant behavior*. Pacific Palisades, CA: Goodyear.

Keith, C. R. (1984). *The aggressive adolescent: Clinical perspectives*. New York: Free Press.

Kupersmidt, J. B., & Patterson, C. J. (1991). Childhood peer rejection, aggression, withdrawal, and perceived competence as predictors of self-reported behavior problems in preadolescence. *Journal of Abnormal Child Psychology*, *19*, 427 – 449.

Kusché, C., & Greenberg, M. (1994). *PATHS: Promoting alternative thinking strategies*. South Deerfield, MA: Developmental Research Programs.

Lewis, H. B. (1971). *Shame and guilt in neurosis*. New York: International Universities Press.

Lewis, M., Alessandri, S. M., & Sullivan, M. W. (1992). Differences in shame and pride as a function of children's gender and task-difficulty. *Child Development*, *63*, 630 – 638.

Lindsay-Hartz, J., De Rivera, J., & Mascolo, M. (1995). Differentiating shame and guilt and their effects on motivation. In J. Tangney & K. Fischer (Eds.), *Self-conscious emotions: Shame, guilt, embarrassment, and pride* (pp. 274 – 300). New York: Guilford.

Lochman, J. E., & Dodge, K. A. (1994). Social cognitive processes of severely violent, moderately aggressive, and nonaggressive boys. *Journal of Consulting and Clinical Psychology*, *62*, 366 – 374.

Loeber, R., & Hay, D. (1997). Key issues in the development of aggression and violence from childhood to early adulthood. *Annual Review of Psychology*, *48*, 371 – 410.

Lord, M. (1999, October 11). The violent kid profile: A controversial new technique for beating violence. *U.S. News & World Report*, *127*, 56 – 57.

Menon, M., Tobin, D. D., Corby, B. C., Menon, M., Hodges, E. V. E., & Perry, D. G. (2007). The developmental costs of high self-esteem. *Child Development*, *78*, 1627 – 1639.

Morf, C. C., & Rhodewalt, F. (2001). Unraveling the paradoxes of narcissism: A dynamic self-regulatory processing model. *Psychological Inquiry*, *12*, 177 – 196.

Morrison, A. P. (1989). *Shame: The underside of narcissism*. Hillsdale, NJ: Analytic.

Neter, J., Wasserman, W., & Kutner, M. H. (1990). *Applied linear statistical models* (3rd ed.). Homewood, IL: Irwin.

Nishina, A., & Juvonen, J. (2005). Daily reports of witnessing and experiencing peer harassment in middle school. *Child Development*, *76*, 435 – 450.

Olthof, T., Ferguson, T. J., Bloemers, E., & Deij, M. (2004). Morality- and identity-related antecedents of children's guilt and shame attributions in events involving physical illness. *Cognition and Emotion*, *18*, 383 – 404.

Olthof, T., Schouten, A., Kuiper, H., Stegge, H., & Jennekens-Schinkel, A. (2000). Shame and guilt in children: Differential situational antecedents and experiential correlates. *British Journal of Developmental Psychology*, *18*, 51 – 64.

Olweus, D. (1994). Bullying at school: Long-term outcomes for the victims and an effective school-based intervention program. In R. Huesmann (Ed.), *Aggressive behavior: Current perspectives* (pp. 97 – 130). New York: Plenum.

Papps, B., & O'Carroll, R. E. (1998). Extremes of self-esteem and narcissism and the experience and expression of anger and aggression. *Aggressive Behavior*, *24*, 421 – 438.

Price, J. M., & Dodge, K. A. (1989). Reactive and proactive aggression in childhood: Relations to peer status and social-context dimensions. *Journal of Abnormal Child Psychology*, *17*, 455 – 471.

Prinstein, M. J., Boergers, J., & Vernberg, E. M. (2001). Overt and relational aggression in adolescents: Social-psychological adjustment of aggressors and victims. *Journal of Clinical Child Psychology*, *30*, 479 – 491.

Prinstein, M. J., & Cillessen, A. H. N. (2003). Forms and functions of adolescent peer aggression associated with high levels of peer status. *Merrill-Palmer Quarterly*, *49*, 310 – 342.

Raskin, R., & Terry, H. (1988). A principal-components analysis of the narcissistic personality inventory and further evidence of its construct validation. *Journal of Personality and Social Psychology*, *54*, 890 – 902.

Reimer, M. S. (1996). "Sinking into the ground": The development and consequences of shame in adolescence. *Developmental Review*, *16*, 321 – 363.

Ringwalt, C. L., Graham, L. A., Paschall, M. J., Flewelling, R. L., & Browne, D. C. (1996). Supporting adolescents with guidance and employment (SAGE). *American Journal of Preventive Medicine*, *12*, 31 – 38.

Robins, R. W., Tracy, J. L., & Shaver, P. R. (2001). Shamed into self-love: Dynamics, roots, and functions of narcissism. *Psychological Inquiry*, *12*, 230 – 236.

Rose, P. (2002). The happy and unhappy faces of narcissism. *Personality and Individual Differences*, *33*, 379 – 391.

Rosenberg, M. (1986). Self-concept from middle childhood through adolescence. In J. Suls & A. Greenwald (Eds.), *Psychological perspective on the self* (Vol. 3, pp. 107 – 135). Hillsdale, NJ: Erlbaum.

Ryan, R. M., & Kuczkowski, R. (1994). The imaginary audience, self-consciousness, and public individuation in adolescence. *Journal of Personality*, *62*, 219 – 238.

Salmivalli, C. (2001). Feeling good about oneself, being bad to others? Remarks on self-esteem, hostility, and aggressive behavior. *Aggression and Violent Behavior*, *6*, 375 – 393.

Salmivalli, C., Kaukiainen, A., Kaistaniemi, L., & Lagerspetz, K. (1999). Self-evaluated self-esteem, peer-evaluated self-esteem, and defensive egotism as predictors of adolescents' participation in bullying situations. *Personality and Social Psychology Bulletin*, *25*, 1268 – 1278.

Simmons, R. G., Rosenberg, F., & Rosenberg, M. (1973). Disturbance in the self-image at adolescence. *American Sociological Review*, *38*, 553 – 568.

Smith, R. H., Webster, J. M., Parrott, W. G., & Eyre, H. L. (2002). The role of public exposure in moral and nonmoral shame and guilt. *Journal of Personality and Social Psychology, 83*, 138 – 159.

Stucke, T. S., & Sporer, S. L. (2002). When a grandiose self-image is threatened: Narcissism and self-concept clarity as predictors of negative emotions and aggression following ego-threat. *Journal of Personality, 70*, 509 – 532.

Swann, W. B., Jr., & Read, S. J. (1981). Self-verification processes: How we sustain our self-conceptions. *Journal of Experimental Social Psychology, 17*, 351 – 372.

Tangney, J. P., & Dearing, R. L. (2002). *Shame and guilt*. New York: Guilford.

Tangney, J. P., Wagner, P. E., Hill-Barlow, D., Marschall, D. E., & Gramzow, R. (1996). Relation of shame and guilt to constructive versus destructive responses to anger across the lifespan. *Journal of Personality and Social Psychology, 70*, 797 – 809.

Thomaes, S., Stegge, H., Bushman, B. J., Olthof, T., & Denissen, J. (2008). Development and validation of the Childhood Narcissism Scale. *Journal of Personality Assessment, 90*, 382 – 391.

Thomaes, S., Stegge, H., & Olthof, T. (2005, August). *"Shame feels angry": On humiliated fury in children.* Poster presented at the 12th European conference on Developmental Psychology, La Laguna, Tenerife, Spain.

Tracy, J. L., & Robins, R. W. (2004). Keeping the self in the self-conscious emotions: Further arguments for a theoretical model. *Psychological Inquiry, 15*, 171 – 177.

Trzesniewski, K. H., Donnellan, M. B., & Robins, R. W. (2003). Stability of self-esteem across the life-span. *Journal of Personality and Social Psychology, 84*, 205 – 220.

Twenge, J. M., & Campbell, W. K. (2003). "Isn't it fun to get the respect that we're going to deserve?" Narcissism, social rejection, and aggression. *Personality and Social Psychology Bulletin, 29*, 261 – 272.

Willock, B. (1987). The devalued (unloved, repugnant) self. A second facet of narcissistic vulnerability in the aggressive, conduct-disordered child. *Psychoanalytic Psychology, 4*, 219 – 240.

Wink, P. (1991). Two faces of narcissism. *Journal of Personality and Social Psychology, 61*, 590 – 597.

Zakriski, A. L., & Coie, J. D. (1996). A comparison of aggressive-rejected and nonaggressive-rejected children's interpretations of self-directed and other-directed rejection. *Child Development, 67*, 1048 – 1070.

Child Development, November/December 2008, Volume 79, Number 6, Pages 1802–1817

Promoting Academic and Social-Emotional School Readiness: The Head Start REDI Program

Karen L. Bierman, Celene E. Domitrovich, Robert L. Nix, Scott D. Gest, Janet A. Welsh,
Mark T. Greenberg, Clancy Blair, Keith E. Nelson, and Sukhdeep Gill
The Pennsylvania State University

Forty-four Head Start classrooms were randomly assigned to enriched intervention (Head Start REDI—*Re*search-based, *D*evelopmentally *I*nformed) or "usual practice" conditions. The intervention involved brief lessons, "hands-on" extension activities, and specific teaching strategies linked empirically with the promotion of: (a) social-emotional competencies and (b) language development and emergent literacy skills. Take-home materials were provided to parents to enhance skill development at home. Multimethod assessments of three hundred and fifty-six 4-year-old children tracked their progress over the course of the 1-year program. Results revealed significant differences favoring children in the enriched intervention classrooms on measures of vocabulary, emergent literacy, emotional understanding, social problem solving, social behavior, and learning engagement. Implications are discussed for developmental models of school readiness and for early educational programs and policies.

Head Start has been referred to as the nation's "premier" federally sponsored early childhood education program developed to reduce socioeconomic disparities in school readiness (U.S. Department of Health and Human Services [USDHHS], 2005). In general, research suggests that comprehensive, high-quality preschool programs can improve the school readiness of disadvantaged children and have extended benefits over time (Barnett, 1995; Ramey & Ramey, 2004). The recent Head Start Impact Study, which compared children randomly assigned to receive Head Start with a similar group allowed to enroll in community non–Head Start services, demonstrated significant benefits. For example, for children attending 1 year of Head Start, statistically significant effects were found on four (of six) cognitive skills measured (with effect sizes ranging from .13 to .34) and on one (of three) social-emotional indices of school readiness (USDHHS, 2005). However, no effects were found on several important aspects of school readiness, including oral comprehension skills, phonological awareness, aggressive behaviors, or social skills. Although Head Start programs provide a strong infrastructure for service delivery, these findings demonstrate a need to further strengthen their educational programs to enhance the impact on child school readiness.

Developed in partnership with Head Start programs, the Head Start REDI (*Re*search-based, *D*evelopmentally *I*nformed) intervention targets the promotion of specific school readiness competencies in the domain of social-emotional development (prosocial behavior, emotional understanding, self-regulation, and aggression control) and cognitive development (language and emergent literacy skills). This article reports on the 1-year child outcomes of REDI assessed using a randomized-controlled design, with multi-informant, multimethod measurement.

Head Start Programs: High/Scope and Creative Curriculum for Preschool

The majority of Head Start programs in the nation use either the High/Scope curriculum, first developed in 1962 (Weikart & Schweinhart, 2005) or the Creative Curriculum for Preschool, first developed in 1978 (Dodge, Colker, & Heroman, 2002). Both of these core curricula emphasize the importance of child-initiated learning supported by positive teacher–child relationships and strategic learning interactions. Brief teacher presentations introduce a mix of large- and small-group exploratory activities each day. In High/Scope, teachers then use a formal "plan, do,

---

This project was supported as part of the Interagency School Readiness Consortium by National Institute of Child Health and Human Development Grants HD046064 and HD43763. We express deep gratitude to Gloria Rhule, Harriet Darling, and Julia Gest for the extensive contributions they made to intervention development and implementation. In addition, appreciation is expressed to the teachers, students, parents, and program personnel who served as partners in this project in the Huntingdon, Blair, and York County Head Start programs of Pennsylvania.

Correspondence concerning this article should be addressed to Karen L. Bierman, The Pennsylvania State University, Department of Psychology, 251 Moore Building, University Park, PA 16802. Electronic mail may be sent to kb2@psu.edu.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7906-0019

review" process to guide children in anticipatory planning, inquiry-guided exploration, and experimentation. Both programs encourage individualized scaffolding (such as "mutual thinking") to elicit student interest and target individualized educational objectives in the context of student activity. Teachers also use large group activities (e.g., games, songs, musical activities) to foster a sense of community and provide practice in social skills and group problem solving. A randomized trial evaluating the High/Scope Perry Preschool Curriculum has demonstrated the inherent potential of this approach, revealing long-term benefits, including higher rates of high school graduation, fewer crimes, and better employment outcomes among children who participated in the Perry Preschool program compared to those who attended no preschool (Schweinhart et al., 2005). Unfortunately, "typical" Head Start programs have not produced the same effects. This is likely due to weaknesses in the fidelity of curriculum implementation and a failure to include the level of parent-focused home visits, teacher training, and organizational support that were used in the High/Scope trial and are recommended by the program developers (Schweinhart, 2004).

Although based on the best research-based developmental practices of the time, both High/Scope and Creative Curriculum originated before contemporary research identified certain predictors of preschool language and emergent literacy skills and before strategies for facilitating social-emotional competencies were empirically evaluated in preschool settings. Both models have incorporated new emphases on literacy and social skills into their programs based on more recent research; however, achieving high-fidelity implementation in the field remains a challenge. Part of the challenge may be that both programs rely on teacher expertise to individualize educational plans. Yet, many Head Start teachers feel overstressed and poorly equipped to handle behaviorally challenging and language delayed children who are often disengaged in the preschool learning context (Yoshikawa & Knitzer, 1997). In addition, Head Start staff often lack the research expertise, financial resources, and technical assistance necessary to integrate new research-based strategies effectively and with high-implementation fidelity, into their local programs (Iutcovich, Fiene, Johnson, Koppel, & Langan, 1997).

Head Start REDI was designed to make the integration of research-based practices easier by providing teachers with manualized enrichment curricula, consisting of brief lessons, "hands-on" extension activities, and specific instructional strategies, all arranged strategically to address a scope and sequence of social-emotional and language/emergent literacy skills. In addition, mentoring was provided to support teacher understanding and effective use of these instructional strategies. The REDI program was designed to enrich and complement the broad educational programming provided by High/Scope or Creative Curriculum, increasing the systematic emphasis teachers placed on target skills. These skills were selected because they are important predictors of grade school adjustment and achievement and are associated with socioeconomic disparities in school readiness (Blair, 2002; Ladd & Profilet, 1996). In addition, in all cases and as described in the following sections, prior efficacy studies had demonstrated instructional approaches that produced improvements in the skills, and these instructional approaches provided a basis for the REDI program design.

Social-Emotional Skills

At school entry, children are faced with heightened demands for well-regulated and goal-directed activity, including sustained behavioral inhibition, compliance with rules, and the capacity to initiate and sustain positive interpersonal relationships with teachers and peers (Campbell & von Stauffenberg, 2008; Kellam, Rebok, Ialongo, & Mayer, 1994). Children growing up in poverty are particularly likely to enter school with significant deficits in social-emotional readiness, with over 40% demonstrating delays in social competencies and communication abilities at school entry and over 20% exhibiting high rates of disruptive behavior problems that undermine school adjustment (Kaiser et al., 2000). To a considerable extent, these child skill deficits reflect the negative impact of poverty—and factors associated with it, such as maternal depression and exposure to violence and stress—on parenting practices (Lengua, Honorado, & Bush, 2007; Li-Grining, 2007).

Target skills. Specific social competencies linked empirically with school success include prosocial behaviors that foster positive peer and teacher relationships (e.g., helping, sharing, taking turns) and self-regulation skills that support the inhibitory control of aggression (Coolahan, Fantuzzo, Mendez, & McDermott, 2000; Denham & Burton, 2003; Ladd, Buhs, & Seid, 2000). Effective prosocial engagement and self-regulation, in turn, appear closely linked with emotional competence (e.g., the ability to recognize and regulate emotions; Denham & Burton, 2003) and with social problem-solving skills (e.g., the capacity to define problems, generate and consider alternative solutions, and engage in anticipatory planning that considers the consequences of various solutions; Crick & Dodge, 1994; Youngstrom et al., 2000).

In addition to these social behaviors, the social-emotional and self-regulation competencies that support effective learning engagement are important for school success (Blair, 2002). These include the capacity to participate cooperatively in classroom activities and to control attention and sustain task involvement (Ladd et al., 2000; McClelland, Acock, & Morrison, 2006). Children who can organize their behavior in a manner consistent with classroom expectations and engage with persistence on learning tasks exhibit higher levels of achievement in school (McClelland et al., 2006). Conversely, attention problems undermine effective learning and contribute to off-task behavior and reduced achievement (Hughes & Kwok, 2006). Although biological factors contribute to individual differences in children's attention skills and effortful control, socialization and educational experiences also appear to play an important role (Blair, 2006). As a result, interventions that foster social-emotional learning and improve behavioral self-regulation can strengthen cognitive development (Riggs, Greenberg, Kusche, & Pentz, 2006).

Intervention approaches. A rapidly growing research base suggests that social-emotional competencies (in areas of prosocial behaviors, aggression control, emotional understanding, social problem-solving skills, and learning engagement) can be enhanced via the use of systematic instructional approaches in the classroom (Consortium on the School-Based Promotion of Social Competence, 1994; Elias et al., 1997). Preschool teachers exert strong socialization influences on young children, affecting social-emotional learning in both formal and informal ways (Denham & Burton, 2003; Pianta & Stuhlman, 2004). Training teachers to provide warm support and effective (nonpunitive) classroom management has positive effects on children's prosocial behavior and reduced aggression (Webster-Stratton, Reid, & Hammond, 2004). In addition, explicit social-emotional learning and emotion coaching curricula enhance child skill development. For example, in a pioneering study, the "I Can Problem Solve" program promoted gains in children's social problem-solving abilities and teacher-rated improvements in frustration tolerance, impulsivity, and task engagement (Shure & Spivack, 1982). Similarly, a randomized trial of the "Al's Pals: Kids Making Healthy Choices" program produced significant effects on Head Start teacher ratings of child behavior problems and independent functioning (Lynch, Geller, & Schmidt, 2004). Further, children with behavior problems who received the "Dinosaur School" program (targeting prosocial skills, emotional understanding, self-regulation, and social problem-solving skills) showed behavioral improvements at home and school that were maintained at follow-up 1 year after the end of the program (Webster-Stratton et al., 2004).

Of particular importance to the current study, a comprehensive, universal social-emotional learning program, Preschool PATHS (Promoting Alternative Thinking Strategies; Domitrovich, Cortes, & Greenberg, 2007) demonstrated positive effects when delivered by teachers in Head Start classrooms. In a randomized trial, Head Start children who received Preschool PATHS showed higher levels of emotional understanding and were rated more socially competent by teachers and parents than children in the control group (Domitrovich et al., 2007).

Language and Emergent Literacy Skills

Language skills support social-emotional adjustment and foster childrens' abilities to understand and comply with the behavioral demands of school (Catts, Fey, Zhang, & Tomblin, 1999). In addition, oral language skills, along with phonological awareness and print knowledge, provide an important foundation for success in the early years of formal reading instruction (Catts et al., 1999; National Research Council, 1998; Stanovich & Siegel, 1994). Children who start school with deficits in these language and emergent literacy skills are frequently identified as poor readers in elementary school and rarely catch up to their nonimpaired peers, often suffering life-long reading disabilities and underachievement (Shaywitz & Shaywitz, 2005).

Target skills. Oral language skills include vocabulary, and the child's ability to understand and produce grammatically varied utterances and narratives. Oral language skills are critical to the process of extracting meaning both from printed text and from teachers' oral instruction. Exposure to a responsive and expansive language environment, in the context of warm, positive relationships with parents and teachers sets the stage for positive language learning (Dickinson & Smith, 1994). Children growing up in families of lower socioeconomic status are less likely than their economically advantaged peers to experience extended discussions with their caregivers or to engage in interactive reading (Senechal & LeFevre, 2002) and hence have less exposure to decontextualized and rich talk that expands vocabulary, fosters narrative skills, and allows them to reflect on their learning experiences.

Phonological awareness, which develops during the preschool years, refers to the ability to recognize and manipulate the smaller units of sound within spoken words, such as syllables and phonemes (Lonigan, Burgess, & Anthony, 2000). Preschoolers who are able to rhyme, blend sounds together to form new sounds, and segment sound into words or syllables are more proficient readers in first and second grades, even when IQ and vocabulary skills are controlled (Bryant, MacLean, Bradley, & Crossland, 1990; Catts et al., 1999). Conversely, children with reading disabilities often exhibit deficits in phonological awareness (Torgesen, 1998). Print knowledge, particularly the child's ability to identify letters of the alphabet, also plays a central role in the child's emerging ability to decode printed text (Scarborough, 2001).

Intervention strategies. Interactive book reading, when combined with professional development activities designed to improve the quality of teacher's language use has proven effective in promoting richer conversational exchanges in the classroom and gains in child vocabulary and oral comprehension skills (Wasik & Bond, 2001; Whitehurst, Arnold, et al., 1994; Whitehurst, Epstein, et al., 1994). In addition, there is compelling evidence that carefully sequenced learning activities can substantially enhance preschool children's phonological awareness and thereby contribute to early reading success (Adams, Foorman, Lundberg, & Beeler, 1998; Hatcher, Hulme, & Ellis, 1994) with some evidence of lasting effects through the end of elementary school (Byrne, Fielding-Barnsley, & Ashley, 2000). In addition, prior efficacy studies have also demonstrated that adding explicit preschool instruction in alphabetic principles fosters later word decoding (Lonigan et al., 2000) and can be combined effectively with phonological awareness interventions to produce gains in reading achievement (Ball & Blachman, 1991; Bradley & Bryant, 1983).

The Present Study

The Head Start REDI program was designed as an enrichment intervention that could be integrated into the existing framework of Head Start programs using the High/Scope or Creative Curriculum. REDI was organized to promote teachers' capacities to use research-based practices in supporting the development of both social-emotional competencies and language/emergent literacy skills. Empirically supported interventions were used or adapted for use in this context, and teachers received training workshops and ongoing weekly mentoring to support their use of the curriculum and teaching strategies. To examine its effects, the REDI Project employed a randomized trial design, assigning 44 classrooms to intervention or "usual practice" conditions. A large, ethnically diverse group of 4-year-old children were pretested as they entered these Head Start classrooms and were assessed again at the end of the year.

It was hypothesized that the REDI intervention would accelerate skill acquisition in the targeted domains of language development (vocabulary, syntax), emergent literacy (phonological awareness, print knowledge), and social-emotional skills (emotional understanding, social problem solving). It was further hypothesized that children would improve in socially competent behavior and learning engagement (increased classroom participation, fewer attention problems) and show reduced aggression, particularly at school, and possibly generalized to home.

Method

Participants

Participants in this study included two cohorts of 4-year-old children (N = 356, 17% Hispanic, 25% African American, and 54% girls) in 44 Head Start classrooms in three counties in Pennsylvania. Half of the participating classrooms came from a large, fairly densely populated county in the southeastern part of the state, which included an urban community surrounded by smaller communities. The other classrooms came from two smaller counties in the central part of the state, characterized by small towns and rural areas. Initially, classrooms were stratified on location, length of program, and student demographics to assure even representation in the intervention and comparison conditions. In the more urban county, classrooms were stratified into three groups, which varied in terms of urban (vs. nonurban) location, percentage of minority students served, and use of Spanish in the classroom. In the two more rural counties, classes were also stratified into three groups, which varied in terms of being full-day or half-day programs and located in small towns versus rural locations. Classrooms in the same center were always assigned to the same condition, to avoid inadvertent contamination of condition within centers. Within each stratified group, centers were randomly assigned to intervention or comparison conditions. This process resulted in 14 classrooms (67%) in each condition that were located in small centers (1–2 classrooms) and 8 classrooms in each condition located in four larger centers (containing 3–5 classrooms), with children from ethnic minority groups fairly evenly spread across condition (39% of the intervention group and 45% of the control group.)

Children were recruited across 2 years. Each year, teachers were studied as they implemented the intervention for the first time, and 4-year-old children were assessed after receiving 1 year of REDI intervention. Teachers in the comparison classrooms continued to conduct curriculum "as usual." On average, classrooms contained about 14 children (typically, one-third 3-year-olds, two-thirds 4-year-olds). Approximately 60% of the 4-year-olds recruited for this study had attended Head Start as 3-year-olds, and this proportion was similar across the intervention and control classrooms. Pretest scores for all children were used to control statistically for initial differences in child skills.

To recruit participants for the evaluation trial, brochures describing the study were distributed to parents of all 4-year-old children in participating classrooms. Overall, only 14 eligible families declined to participate in the evaluation procedures, but an additional 21 primary caregivers were very difficult to reach and failed to complete the preassessment. Two children were dropped because they had a twin or sibling in the study, and 19 families withdrew early from Head Start and hence completed only parts of the assessment procedures. Hence, study participants represent 86% of the initially eligible population. Primary caregivers included 89% mothers, 4% fathers, 4% grandparents, and 3% other, including stepparents or foster parents.

Data Collection Procedures

Parent interviews were conducted at a time and location that was convenient for the parents (e.g., at the Head Start center, their home, or at a community location such as a library). A trained interviewer met privately with the primary caregiver and read through all the measures. Based on teacher recommendation or parent request, some interviews were conducted in Spanish. Parent interviews were initiated shortly before the start of the school year and continued through the end of October. Parents were reinterviewed at the end of the school year, in May and June. Each time, parents were compensated $20 for this 30-min interview.

Child assessments were conducted at school by trained interviewers, during two individual 30- to 45-min "pull-out" sessions. After each session, interviewers rated the child's task-approach and learning behaviors, as observed during the session. To give children time to acclimate to the classroom setting, assessments began 3 weeks after school began and continued through the end of October. End-of-year child assessments were conducted in March and April.

End-of-year child ratings were collected from teachers in April. Teachers were compensated $20 to provide general information about themselves and their classrooms, and they received an additional $7 per child for completing behavioral ratings. One lead and one assistant teacher in each classroom provided independent ratings of child behavior.

Beginning in April, each child was observed during two 12- to 15-min play sessions held on 2 separate days. Using a roster generated by the research staff, classmates were randomly assigned to groups of three, taken to a separate space, and given the opportunity to play with an interesting and novel toy together (playground set and castle set). By using a standard play setting that encouraged social interaction, these sessions were designed to control for classroom variations in play opportunities and to elicit meaningful individual differences in social competencies. Observers were naive concerning the intervention or control status of the children. One observer focused on each child, noting discrete social behaviors and interactions, and at the end of the session, completed rating forms describing the quality of social behavior observed.

Intervention Design

The intervention was delivered by classroom teachers and integrated into their ongoing classroom programs. It included curriculum-based lessons, center-based extension activities, and training in "teaching strategies" to use throughout the day.

Social-emotional skill enrichment. The Preschool PATHS Curriculum (Domitrovich, Greenberg, Cortes, & Kusche, 1999) was used to promote children's social-emotional skills. This curriculum targets four domains: (a) prosocial friendship skills, (b) emotional understanding and emotional expression skills, (c) self-control (e.g., the capacity to inhibit impulsive behavior and organize goal-directed activity), and (d) problem-solving skills, including interpersonal negotiation and conflict resolution skills. The curriculum is divided into 33 lessons that are delivered by teachers during circle time. These lessons include modeling stories and discussions and use puppet characters, photographs, and teacher role-play demonstrations. Each lesson includes extension activities (e.g., cooperative projects and games) that provide children with opportunities to practice the target skills with teacher support. Teachers taught one PATHS lesson and conducted one extension activity each week.

Generalized teaching strategies were encouraged with mentoring, including positive classroom management, use of specific teacher praise and support, emotion coaching, and induction strategies to promote appropriate self-control.

Language/emergent literacy skill enrichment. Four language and emergent literacy skills were targeted in REDI: (a) vocabulary, (b) syntax, (c) phonological awareness, and (d) print awareness. Three program components were developed to target these skills, including an interactive reading program, a set of "Sound Games," and print center activities.

The interactive reading program was based on the shared reading program developed by Wasik and Bond (2001; Wasik, Bond, & Hindman, 2006), which was, in turn, an adaptation of the dialogic reading program (Whitehurst, Arnold, et al., 1994). The curriculum included two books per week, which were scripted with interactive questions. Each book had a list of targeted vocabulary words, presented with the aid of physical props and illustrations. In addition to presenting these materials in a systematic way during the week, teachers received mentoring in the use of "language coaching" strategies, such as expansions and grammatical recasts, to provide a general scaffold for language development in the classroom (Dickinson & Smith, 1994). The overall goal was to improve teacher's strategic use of language in ways that would increase child oral language skills, including vocabulary, narrative, and syntax.

Teachers were provided with curricula materials to promote phonological awareness and print knowledge. A set of Sound Games was based primarily on the work of Lundberg and colleagues (Adams et al., 1998). The games were organized developmentally, moving from easier to more difficult skills during the course of the year (e.g., listening, rhyming, alliteration, words and sentences, syllables, phonemes.) Teachers were asked to use a 10- to 15-min Sound Game activity at least three times per week.

In addition, teachers were provided with a developmentally sequenced set of activities and materials to be used in their alphabet centers, including letter stickers, a letter bucket, materials to create a Letter Wall, and craft materials for various letter-learning activities. They were asked to make sure that each child visited the alphabet center several times per week and were given materials to track the children's acquisition of letter names.

Training and professional development support. Teachers received detailed manuals and kits containing all materials needed to implement the intervention. A 3-day professional training was conducted in August, prior to initiating the intervention, and a 1-day "booster" training session was conducted in January. Teachers also received weekly mentoring support provided by local educational consultants ("REDI trainers"), experienced master teachers who were supervised by two project-based senior educational trainers. The weekly consultations were intended to enhance the quality of implementation through modeling, coaching, and providing ongoing feedback regarding program delivery. REDI trainers spent an average of 3 hr per week ($SD = 0.18$) in each classroom observing, demonstrating, or team teaching lessons. They also met with the head and assistant teacher for 1 hr each week outside of class.

Parent take-home materials. Three "take-home" packets were mailed to parents during the course of the year, each containing a modeling videotape, with parenting tips and learning activities to use at home. In addition, the PATHS curriculum included handouts for parents, with suggestions for home activities. Children also took home letter stickers and compliment pages to prompt their parents to ask them about their school day and provide positive support at home.

Integrating the program components. Half of the participating classrooms used High/Scope as their base curriculum; the others used Creative Curriculum. Project staff developed "crosswalk" tables to illustrate how the REDI target skills and methods mapped onto each of these base curricula. Although both High/Scope and Creative Curriculum have recently developed supplemental materials to enhance language and emergent literacy skills, none of the Head Start programs participating in this project were using them. The interactive reading program and emphasis on language use, as well as the emotion coaching and social problem-solving strategies that were central to the REDI program were philosophically compatible with the strategic, child-centered teaching approach used in High/Scope and Creative Curriculum. They were easily integrated into existing book reading, learning center, and play activities and routines. The Sound Games, alphabet centers, and social-emotional lesson presentations used more direct teacher instruction, and hence represented components that were not well integrated with the core curriculum but were "added on" to enhance the acquisition of specific skill concepts.

A central objective of REDI was to maximize the integration of the social-emotional and language/emergent literacy intervention components that comprised the enrichment program. Each week, one of the books used in the interactive reading program focused on the PATHS theme for that week (e.g., friendship, feelings, self-control, social problem solving), and feeling words were included in the vocabulary prompts. Conversely, PATHS extension activities incorporated language and emergent literacy skills.

Intervention implementation. Teachers reported to REDI trainers regarding the enrichment lessons and activities they used. On average, each week they reported implementing 1.77 PATHS lessons and extension activities (*SD* = 0.12), 6.08 dialogic reading activities (*SD* = 0.25), 2.57 sound game activities (*SD* = 0.34), and 3.56 alphabet center activities (*SD* = 0.18). In addition, teachers answered 10 questions, using 3-point scales to describe the quality of their implementation (e.g., Were you able to complete the lesson as written? How well did the children understand the major points of the lesson?; α = .87). The average rating was 2.78 (*SD* = 0.15) indicating that, from the teachers' perspective, the curriculum was being delivered with high fidelity, and children were engaged in the lessons.

On a monthly basis, REDI trainers assessed the fidelity and quality of implementation of program components, using a 6-point Likert scale (*poor* to *exemplary*). Implementation quality ratings were: PATHS, *M* = 4.61 (*SD* = 0.74), dialogic reading, *M*=4.39 (*SD* = 0.57), alphabet activities, *M* = 4.70 (*SD* = 0.55), sound games, *M*=4.52 (*SD* = 0.72), and overall REDI program, *M* = 4.55 (*SD* = 0.67). Scores of 4–5 reflected descriptions of "adequate" to "strong."

Measures

A multimethod, multimeasure assessment battery included child assessments, teacher ratings, parent ratings, and direct observations. Outcome measures represented six core domains: (a) language skills, (b) emergent literacy skills, (c) emotional understanding and social-cognitive skills, (d) social behaviors, (e) learning engagement at school, and (f) learning engagement at home.

Language skills. Three tests were administered directly to children to assess their language skills. In the Expressive One-Word Picture Vocabulary Test (Brownell, 2000), children gave the word that best described pictures they were shown (α = .94.) The Grammatical Understanding subtest of the Test of Language Development (TOLD; Newcomer & Hammill, 1997) assessed syntax comprehension. Children listened to a sentence and chose one of the four pictures that "best matched" the meaning of the sentence (α = .80). The Sentence Imitation subtest assessed syntax expression. Children repeated sentences read aloud by the interviewer. Scores reflected the number of increasingly complex sentences a child imitated correctly (α = .90). Prior research has documented strong test–retest reliability for TOLD subtests (*r* = .90) and concurrent validity with comprehensive language assessments (Newcomer & Hammill, 1997).

Emergent literacy skills. Three subscales assessing emergent literacy skills were drawn from the Test of Preschool Early Literacy (TOPEL; previously labeled the Pre-CTOPP; Lonigan, Wagner, Torgesen, & Rashotte, 2007). The Blending subtest assessed phonological processing. Children were asked to combine different parts of a word, such as "hot" and "dog" or "b" and "air" and point to the correct picture or say the full word (α = .86). On the Elision subtest of the TOPEL, children deconstructed compound words and pointed to the correct picture (e.g., Point to "snowshoe" without "snow"; Say "airport" without "air"; α = .83). In the Print Knowledge subtest of the TOPEL, children identified pictures of letters or words and named letters (α = .97). Prior research has reported correlations in the range of .43–.88 between these three subscales and the acquisition of initial reading skills (Lonigan, 2006).

Emotional understanding and social-cognitive skills. Two measures assessed emotional understanding. On the Assessment of Children's Emotion Skills (Schultz, Izard, & Bear, 2004), children determined whether the facial expressions in 12 photographs reflected happy, mad, sad, scared, or no feelings. The score was the total number correctly identified (α = .57). On the Emotion Recognition Questionnaire (Ribordy, Camras, Stafani, & Spacarelli, 1988), children listened to 16 stories describing characters in emotionally evocative situations and identified their feeling by pointing to pictures of happy, mad, sad, or scared faces. Children received a score of 2 for correctly identifying the feeling and a score of 1 for correctly identifying the valence (α = .63).

Social problem-solving skills were assessed using a variation of the Challenging Situations Task (Denham, Bouril, & Belouad, 1994.) Children were presented with pictures of four peer scenarios (e.g., a peer knocking down blocks, being hit, entering a group, a peer taking a ball). After each scenario, children were asked what they would do in the situation. Their open-ended responses were coded as competent (i.e., appropriately asserting oneself or calmly negotiating a solution, α = .68), aggressive (i.e., responding with verbal or physical antagonism, intimidation, or force, α = .77), or inept (i.e., passive avoidance, α = .68). Coding agreement between research assistants and data checkers on 1,616 responses was high (κ = .94).

Social-emotional behaviors. Teacher ratings, observer ratings, and parent ratings assessed social competence and aggressive-oppositional behavior. The 13 items of the Social Competence Scale (Conduct Problems Prevention Research Group [CPPRG], 1995) were rated on a 6-point Likert scale (*never* to *almost always*) and included prosocial behaviors such as sharing, helping, understanding other's feelings, as well as self-regulatory behaviors, such as resolving peer problems independently. Ratings provided by lead and assistant teachers were averaged (r = .56, p < .001). Internal consistency was high for teachers (α = .94) and parents (α = .87). Observers used the same rating scale to describe child behavior after each of the play observation sessions (α = .88). Interrater reliability was assessed for 23% of the playgroup sessions and demonstrated adequate agreement (intraclass correlation coefficient [ICC] = .70). Ratings collected after each of the two playgroup sessions were averaged (r = .24, p < .001).

Seven items from the Teacher Observation of Child Adaptation – Revised (TOCA – R; Werthamer-Larsson, Kellam, & Wheeler, 1991) assessed overt aggression (e.g., stubborn, yells, fights). Six items from the Preschool Social Behavior Scale – Teacher Form (Crick, Casas, & Mosher, 1997) assessed relational aggression (e.g., "Tells other kids he/she won't be their friend unless they do what he/she wants"). Items were rated on a 6-point Likert scale (*almost never* to *almost always*; α = .88 and .93, for the two scales, respectively). Ratings from lead and assistant teachers were averaged (r = .68, p < .001, for overt aggression, and r = .51, p < .001, for relational aggression), and overt and relational ratings were combined to form a total aggression score (r = .58, p < .001). Parents and observers completed the seven items from the TOCA – R only (α = .86, α = .92, respectively). Interrater agreement among observers was acceptable (ICC = .74), and ratings were averaged across the two sessions (r = .42, p < .001).

Learning engagement at school. Teacher ratings were used to assess learning engagement at school, using an eight-item inventory developed for this study. Items were rated on a 6-point Likert scale (*strongly disagree* to *strongly agree*), and reflected self-regulation (e.g., "Has the self-control to do well in school" and "Can follow the rules and routines that are part of the school day"), learning motivation and involvement (e.g., "Seems enthusiastic about learning new things"), and compliance (e.g., "Is able and willing to follow teacher directions"). Lead and assistant teacher ratings were averaged (α = .95, r = .70, p < .001).

Observers completed the Adapted Leiter – R Assessor Report (Roid & Miller, 1997) to assess task orientation at school (Smith-Donald, Raver, Hayes, & Richardson, 2007). After the child assessments, interviewers rated the child's participation and involvement, using 13 items (e.g., "Shows pleasure in accomplishment and active mastery," "Careful and interested in accuracy," "Alert and interactive.") Each item was rated on a 4-point scale (α = .93). Scores from the two assessment sessions were averaged (r = .62, p < .001).

Teachers also completed the ADHD Rating Scale (DuPaul, 1991). Based on the *Diagnostic and Statistical Manual* symptoms of attention deficit hyperactivity disorder, this scale includes 14 items reflecting attention problems (e.g., "Is easily distracted," "Has trouble following directions") each rated on a 4-point scale. Lead and assistant teacher ratings were averaged (α = .94; r = .76, p < .001) and used in this study to represent a continuous dimension reflecting difficulties with impulse control, distractibility, and sustained attention.

Learning engagement at home. Parents completed the ADHD Rating Scale (α = .90). In addition, they responded to five questions about children's language and communication at home (e.g., "How many times in a typical week do you and your child have a conversation that lasts 10 min or more?" "How often does your child volunteer to tell you about something that happened when you were not with him or her?"; α = .49). Finally, parents answered six questions about children's engagement in reading at home (e.g., "When was the last time you and your child read a book together?" "How many books did you read at that time?"; α = .63).

Results

Preliminary Analyses

Scores on Sentence Imitation and observer ratings of aggressive-oppositional behavior were highly skewed. Log transformations were conducted on those scores, after which all measures were normally distributed with values of skewness and kurtosis below 2.00. Correlations among the measures of child skills (language, emergent literacy, and social-emotional cognitions) are presented in Table 1, and correlations among the measures of child behavior ratings (by teachers, parents, and observers) are presented in Table 2.

To examine preintervention differences between the intervention and the comparison groups, hierarchical linear models (HLM) were estimated, accounting for the nesting of children within classrooms. Child sex and race were included as Level 1 covariates

Table 1
Correlations Among Child Skills

	1	2	3	4	5	6	7	8	9	10	11
1. Vocabulary											
2. Grammatical understanding	.50										
3. Sentence imitation	.47	.39									
4. Blending	.35	.30	.40								
5. Elision	.52	.43	.45	.50							
6. Print awareness	.48	.40	.37	.37	.45						
7. Emotion identification (Assessment of Children's Emotion Skills)	.38	.43	.33	.32	.29	.33					
8. Emotion recognition (Emotion Recognition Questionnaire)	.44	.45	.39	.35	.38	.40	.42				
9. CST competent responses	.18	.21	.14	.17	.13	.19	.12	.17			
10. CST aggressive responses	−.06	−.06	−.01	−.07	−.06	−.03	−.09	−.09	−.37		
11. CST inept responses	−.21	−.21	−.18	−.14	−.26	−.17	−.12	−.20	−.27	−.16	

Note. All correlation coefficients greater than .10 are statistically significant ($p < .05$). CST = Challenging Situations Task.

and center site (e.g., central or southeastern Pennsylvania), cohort, and intervention status were included as Level 2 covariates in each of these models. No significant ($p < .05$) pretreatment intervention-control group differences emerged on the 11 measures of child skills or the 6 measures of child behavior for which pretreatment scores were available. Overall, the stratified randomization process appeared effective in creating equivalent groups prior to the intervention.

Postintervention Group Differences

The first set of analyses examined the 11 measures of child language, emergent literacy, emotional understanding, and social problem-solving skills that were targeted directly by the intervention and were measured using direct assessments of child skill knowledge. Because the intervention was delivered at the classroom level, analyses accounted for the nonindependence of data at that level. Child sex and race were included as Level 1 covariates and site, cohort, and intervention status were included as Level 2 covariates. Preintervention scores were available for all these measures and also were included as Level 1 covariates.

The results of these HLM are summarized in Table 3. For ease of interpretation, all measures were standardized with a mean of 0 and a standard deviation of 1. By doing this, the coefficient for the intervention effect represents the difference in average expected scores between children in the intervention and control groups as a proportion of a standard deviation. This difference is similar to

Table 2
Correlations Among Behavioral Ratings

	1	2	3	4	5	6	7	8	9	10	11
1. Social competence: Teacher											
2. Social competence: Parent	.18										
3. Social competence: Observer	.29	.18									
4. Aggression: Teacher	−.67	−.09	−.26								
5. Aggression: Parent	−.26	−.70	−.13	.22							
6. Aggression: Observer	−.15	−.11	−.62	.23	.13						
7. Learning engagement: Teacher	.73	.22	.26	−.39	−.17	−.11					
8. Task orientation: Observer	.28	.15	.28	−.10	−.13	−.26	.41				
9. Attention problems: Teacher	−.72	−.14	−.35	.60	.19	.28	−.73	−.32			
10. Attention problems: Parent	−.21	−.49	−.22	.20	.54	.10	−.16	−.20	.21		
11. Language use/communication	−.05	.14	.02	.08	.02	.05	.01	.05	.06	−.01	
12. Reading activities	−.02	.13	−.04	.08	−.07	.06	.07	.09	.02	−.11	.35

Note. All correlation coefficients greater than .10 are statistically significant ($p < .05$).

Table 3
Intervention Effects on Child Skills

Variable	Control group		Intervention group		Intervention effect	
	Pre	Post	Pre	Post	Effect	*p* Value
Language skills						
Vocabulary	34.62 (10.76)	41.03 (11.24)	34.94 (11.10)	42.79 (11.55)	.15	.02
Grammatical understanding	8.62 (5.14)	11.72 (4.90)	9.72 (4.95)	11.95 (4.70)	−.07	.56
Sentence imitation	1.11 (0.72)	1.56 (0.80)	1.19 (0.79)	1.57 (0.75)	−.04	.73
Emergent literacy skills						
Blending	11.53 (4.30)	13.04 (4.30)	11.58 (4.26)	14.71 (4.33)	.39	.001
Elision	7.71 (3.43)	9.61 (3.58)	8.21 (3.49)	11.19 (4.02)	.35	.001
Print awareness	7.97 (8.31)	16.49 (12.59)	8.86 (9.43)	18.84 (12.84)	.16	.10
Social cognitions						
Emotion identification	5.68 (2.27)	7.12 (2.28)	5.39 (2.16)	7.45 (2.36)	.21	.06
Emotion recognition	1.34 (0.29)	1.52 (0.26)	1.37 (0.31)	1.61 (0.24)	.23	.03
CST aggressive responses	2.01 (2.30)	2.06 (2.55)	1.96 (2.27)	1.53 (2.21)	−.21	.04
CST competent responses	2.37 (2.05)	2.29 (2.05)	2.51 (2.16)	3.16 (2.55)	.35	.005
CST inept responses	2.52 (2.11)	2.56 (2.34)	2.51 (2.11)	1.94 (2.01)	−.28	.03

Note. Group means are presented with standard deviations in parentheses. Effect estimates represent approximate effect size. CST = Challenging Situations Task.

an effect size that controls for all other measures in the HLM. There were significant treatment effects ($p < .05$) for 7 of the 11 skills directly targeted by the intervention and marginally significant intervention effects ($p < .10$) for 2 other skills. Only 2 of the skills (grammatical understanding and sentence imitation) showed no evidence of intervention impact. The effect size for the 9 skills showing an intervention effect ranged from .15 to .39. This suggests that children who received the REDI enrichment gained about one fourth of a standard deviation more in these areas of target skill acquisition than children in the control group who received usual practice Head Start.

Next, HLM analyses were conducted to assess intervention effects on the 12 behavioral ratings provided by teachers, observers, and parents. Ideally, the cognitive and social-emotional skills acquired by children in the intervention group, along with the emotion coaching and inductive support provided by their teachers should produce generalized improvements in social competence, aggression control, and learning engagement. As before, the HLM nested children within classroom, included child sex and race as Level 1 covariates and included site, cohort, and intervention status as Level 2 covariates. In this case, however, we only had preintervention parent ratings and preintervention observer ratings of task orientation as Level 1 covariates. As shown in Table 4, 3 of the 12 behavioral ratings showed significant ($p < .05$) intervention effects—teacher ratings of aggression, observer ratings of task orientation, and parent ratings of communication/language use. Five additional measures showed nonsignificant trends ($p < .10$) favoring intervention—teacher ratings of social competence, observer ratings of social competence, and parent ratings of aggression, attention problems, and reading involvement. Only four measures (parent ratings of social competence, observer ratings of aggression, teacher ratings of learning engagement, and teacher ratings of attention problems) showed no evidence of intervention impact. On the dimensions that showed marginal or statistically significant intervention effects, the effect sizes ranged from .11 to .29. Hence, children who received the REDI enrichment showed behavioral improvements in these areas that averaged about one fifth of a standard deviation more than the behavioral maturation observed in control group children.

Discussion

The Head Start REDI intervention was designed to capitalize on the existing service delivery infrastructure and quality programming of the Head Start system. The goal was to improve program impact on child skill acquisition by increasing the teachers' use of research-based instructional materials and teaching strategies. To do so, REDI provided supplementary curricula and mentoring to assist teachers

Table 4
Intervention Effects on Child Behaviors

	Control group		Intervention group		Intervention effect	
Variable	Pre	Post	Pre	Post	Effect	*p* value
Social-emotional behaviors						
Social competence: Teacher	NA	3.98 (0.88)	NA	4.15 (0.82)	.24	.10
Social competence: Parent	3.59 (0.76)	3.66 (0.84)	3.60 (0.84)	3.73 (0.80)	.09	.24
Social competence: Observer	NA	2.21 (0.53)	NA	2.36 (0.49)	.26	.08
Aggression: Teacher	NA	4.12 (1.70)	NA	3.69 (1.52)	−.28	.05
Aggression: Parent	2.93 (0.94)	2.86 (0.97)	2.88 (1.02)	2.71 (0.99)	−.13	.08
Aggression: Observer	NA	0.37 (0.34)	NA	0.30 (0.31)	−.19	.23
Learning engagement at school						
Learning engagement: Teacher	NA	4.97 (0.90)	NA	5.08 (0.86)	.17	.11
Task orientation: Observer	3.35 (0.48)	3.38 (0.49)	3.27 (0.60)	3.49 (0.49)	.29	.02
Attention problems: Teacher	NA	0.73 (0.66)	NA	0.66 (0.63)	−.14	.21
Learning engagement at home						
Attention problems: Parent	1.28 (0.62)	1.64 (0.70)	1.24 (0.68)	1.48 (0.81)	−.10	.06
Language/communication: Parent	0.00 (0.59)	−0.07 (0.52)	0.00 (0.55)	0.06 (0.56)	.25	.02
Reading activities: Parent	−0.01 (0.63)	−0.05 (0.65)	0.01 (0.58)	0.05 (0.51)	.17	.07

who were using High/Scope or Creative Curriculum to incorporate language more effectively throughout the day, provide both explicitly taught and more generalized support for child social-emotional development, and include explicit instruction in phonological sensitivity and letter identification skills.

Prior studies have shown that model preschool programs can have strong and long-lasting effects on the educational attainment and life adjustment of children growing up in socioeconomically disadvantaged situations. However, Head Start programs are often underresourced in comparison, and many lack the high levels of teacher training and organizational support that characterize the most successful models (Schweinhart, 2004). Both High/Scope and Creative Curriculum encourage teachers to promote child-centered learning, individualizing their approach to expand on individual student interests and strengths. In addition, teachers are expected to use professional development workshops to master new, research-based strategies for promoting emergent literacy and social-emotional competences and to incorporate these strategies into their classroom practice. These expectations can create "overload" for teachers, who often lack the time, technical assistance, financial resources, and organizational support needed for effective and high-fidelity implementation (Iutcovich et al., 1997; Yoshikawa & Knitzer, 1997).

Head Start REDI sought to support teachers and improve the fidelity of research-based instructional practice in Head Start classrooms, by providing teachers with a well-specified enrichment curriculum, along with a year of mentored professional development in the implementation of that curriculum and corresponding teaching practices. The curriculum included specific instructional strategies and hands-on extension activities that were strategically diffused within the context of regular High/Scope and Creative Curriculum activities, as well as "scripted" lessons, all carefully organized to address a scope and sequence of language/emergent literacy and social-emotional skills.

The Program Impact on Child School Readiness

Language skills and social-emotional competencies were targeted for emphasis in REDI because these skills provide fundamental support for effective school engagement; they facilitate the child's ability to follow classroom rules, cope actively with learning challenges, and relate to teachers and peers (McClelland et al., 2006). These skills are intertwined developmentally. Language skills enhance the child's capacity to regulate emotions and promote effective social interaction (Greenberg, Kusche, & Speltz, 1991), and conversely, social-emotional competencies foster positive adult and peer relationships, motivating and providing important opportunities for language learning and cognitive development (Bierman, Greenberg, & the Conduct Problems Prevention Research Group, 1996). A key feature of the REDI program was the synchronized and integrated focus on both language/emergent literacy skills and social-emotional competencies, designed to support skill development in both domains and foster transactional processes promoting cross-domain and comprehensive school readiness.

Teachers taught explicit lessons on social-emotional skills and, with mentoring, were encouraged to use emotion coaching, induction strategies, and problem-solving dialogue techniques throughout the day to encourage child social-emotional development. Statistically significant effects emerged on direct assessments of children's emotional understanding and social problem-solving skills, reflecting modest effect sizes (range = .21 – .35). Corresponding improvements were also noted on measures of behavioral functioning, although not as consistent or strong as the impact on the targeted social-cognitive skills. Specifically, statistically significant differences in aggression were reported by teachers and statistically significant improvements in task orientation were reported by observers. Nonsignificant trends also favored children in the intervention classrooms on measures of teacher- and observer-rated social competence and parent-rated aggression and attention problems. Effect sizes ranged from .11 to .29. These are relatively small effects and require follow-up assessment in elementary school to determine whether they promote generalized and sustained improvements in school adaptation over time.

Additional REDI program components focused on improving classroom support for child language development (Dickinson & Smith, 1994). The interactive reading program provided teachers with scripted books and targeted vocabulary. Mentors encouraged teachers to elicit children's language more effectively and respond to it more sensitively. Mentors also encouraged teachers to model decontextualized and rich talk and to use questions and summaries to foster child narrative skills (Wasik & Bond, 2001; Whitehurst, Arnold, et al., 1994). Significant effects favoring children in intervention classrooms emerged on child vocabulary and on parent reports of communication and language use at home (with effect sizes of .15 and .25, respectively), but no effects emerged on the measures of child grammatical understanding.

In addition, the REDI curriculum used explicit instruction to promote the discrete skills of phonological awareness and print knowledge (Lonigan et al., 2000; Scarborough, 2001). Two indices of phonological awareness (blending and elision) showed a significant intervention effect; print awareness showed a nonsignificant trend favoring intervention (effect sizes ranged from .16 to .39). These findings suggest that the intervention successfully promoted the discrete skills targeted in this domain. However, follow-up data are required to determine whether these discrete skills foster early reading skill acquisition, as they are hypothesized to do.

Several major components of the REDI program—specifically, sensitive-responsive language use, emotion coaching, and social problem-solving dialogue—were philosophically compatible with the strategic, child-centered teaching approach used in High/Scope and Creative Curriculum. However, the use of explicit instruction (PATHS lessons, sound games, print center focus on letter identification) was less consistent with this philosophy because of the emphasis on direct instruction. Proponents of "developmentally appropriate practice" in preschool settings have concerns about explicit instructional techniques of this kind that teach skills in a decontextualized fashion. The main concern is that an overemphasis on explicit instruction may reduce the learning engagement of young children and stifle curiosity and exploratory learning. Countering these concerns are proponents of explicit instruction who cite evidence that emergent literacy skills (such as letter identification and phonemic awareness) are strong predictors of later school achievement that are not likely to be acquired by children in Head Start unless explicit instruction techniques are employed (Lonigan et al., 2000). The results of the Head Start REDI program suggest that it is possible to integrate empirically validated strategies for promoting these critical emergent literacy skills in ways that are consistent with Head Start practices. With support and mentoring, Head Start teachers implemented the REDI curriculum components with a high level of fidelity and quality throughout the year. Interactive reading techniques were integrated into preexisting read-aloud sessions, alphabet center activities were integrated into preexisting print centers, and PATHS lessons were conducted during preexisting morning "circle times." Sound games were most often added during transition periods, and PATHS extension support (emotion coaching and social problem-solving support) was used during free play periods. The significant effects on child skill acquisition suggest that this hybrid approach of incorporating some explicit instruction in the context of a child-centered, comprehensive early childhood program may be beneficial. Follow-up assessments are needed to determine the sustainability of these gains as well as to determine whether the program had any costs for children, in terms of reduced motivation or decreased gains in other domains of development.

Strengths and Limitations of the Study

Several strengths of the study warrant mention. The randomized-controlled design and use of multiple informants enhance confidence in the validity of the findings. Although teachers knew which children were receiving intervention, child assessors, observers, and parents were naive concerning the intervention/control status of the children. Hence, most of the measures used in the study were unbiased. HLM were used to control for the nesting of children within classrooms and to adjust for the potential dependencies in intervention effects associated with the classroom implementation of the program.

An added feature of this study involved the use of usual practice Head Start classrooms for a control group. Many of the studies that have produced larger effect sizes for preschool interventions have relied on comparison children who received no or minimum preschool support. This study demonstrated the effects of the enrichment relative to Head Start classrooms using standard High/Scope or Creative Curriculum, increasing the confidence with which the findings can be attributed to the specific teacher mentoring and supplementary curriculum used, rather than to more general features associated with the Head Start preschool experience.

A limitation of the study that should be noted, however, is that no measures were collected to determine the fidelity with which these programs were implementing their basic High/Scope or Creative Curriculum. It is possible that, had an equal amount of professional development effort been exerted to improve and monitor the fidelity of these core curricula, the results would have been similar to the impact of adding new curricular components. We hypothesize that providing teachers with a well-specified curriculum and lesson plans that are ordered along a developmental sequence makes it easier for them to introduce concepts and learning activities in a systematic and strategic manner. However, similar to most models of early childhood education, the REDI program assumes that young children acquire new skills when they are actively interested and engaged in goal-oriented activity, and hence that effective teaching requires sensitive and flexible responding to child-initiated inquiries and responses. For this reason, the REDI program included both a well-specified curriculum and a heavy emphasis on professional development and mentoring for teachers. The consequence is that one cannot determine whether either one of these components (without the other) would have been sufficient to promote the same child outcomes. It is quite possible that changes in teaching practices (e.g., enriched language use, more positive behavioral support, improved instructional quality) were the key features that accounted for positive changes in child outcomes. The REDI curriculum and mentoring program are likely just one of the several approaches that improve child outcomes via improvements in teaching quality.

In addition, the parent component of the REDI intervention was minimal, consisting of take-home handouts, videos, and a few parent–child activities. Even so, parents reported significant differences in child communication and language use at home and nonsignificant trends favoring intervention on the child's reading involvement, aggression, and attention problems at home. To a large degree, the negative impact of family poverty on child school readiness is mediated by home learning opportunities, particularly the quality of parent–child interactions and associated provision of cognitive stimulation and emotional support (Campbell & von Stauffenberg, 2008; Duncan, Brooks-Gunn, & Klebanov, 1994; Eccles & Harold, 1996). Hence, it seems quite likely that greater attention to parenting might strengthen program impact.

A third limitation of the present study involves the timing of the assessments. Because we needed to give children some time to acclimate to the classroom setting, and then work around classroom schedules, it took us 8 weeks at the beginning of the year to conduct all the child preintervention assessments. Similarly, we had to start the child postintervention assessments 8 weeks before the end of the school year. On average, the impact on child skills was evaluated on the basis of approximately 25 weeks of exposure of the 35-week program year.

Fourth, because we needed to assess young children with limited attention on a wide range of cognitive and social-emotional domains, we had to rely on very brief assessments of discrete skills that are assumed to reflect broader domains of school readiness. We do not know, for example, how well some of our indicators of emergent literacy will predict future reading ability, and we do not know how well postintervention differences on teacher ratings or the relatively brief behavioral observations will predict later social skills.

Relatedly, although REDI appears to have had a positive effect on children's development, we cannot know how robust those changes are. Follow-up data are needed on these children as they enter kindergarten and proceed into first grade to determine whether the gains observed on these measures will predict longer range benefits in school achievement and behavioral adjustment.

Implications for Early Educational Policies and Programs and Future Directions

The present findings suggest that the impact of Head Start on children's cognitive and social-emotional school readiness can be improved by incorporating research-based instructional strategies and teaching practices into existing high-quality curricula, such as High/Scope and Creative Curriculum. They also suggest that usual practice strategies for teacher training in research-based practices are insufficient. That is, during the course of this trial, teachers in control classrooms participated in Head Start – sponsored professional development workshops but were not able to leverage this training into changes in classroom practice that supported the same rate of student skill acquisition as did REDI. The specific and scripted REDI curricula may provide critical scaffolds to enhance improved practice, assisting teachers by organizing skill presentation along a scope and sequence, reducing teacher preparation time, and providing teachers with a platform for skill coaching throughout the day. This type of intensive skill focus may be most effective when it is integrated into a comprehensive, child-centered educational program, such as High/Scope and Creative Curriculum, which provides broad support for the child's overall development and emphasizes emotionally supportive and responsive teaching.

More broadly, the present program validates the strategy of enriching current "best practices" curricula with emerging research-based curriculum materials and teaching strategies. Given the limited number of hours in the school day, teachers sometimes feel that they must choose between focusing on the cognitive skills or the social-emotional needs of socioeconomically disadvantaged students. These findings demonstrate that a dual-focus integrated intervention model can effectively and simultaneously promote gains in both domains.

References

Adams, M. J., Foorman, B. R., Lundberg, I., & Beeler, T. (1998). *Phonological sensitivity in young children: A classroom curriculum*. Baltimore: Brookes.

Ball, W. W., & Blachman, B. A. (1991). Does phoneme segmentation training in kindergarten make a difference in early word recognition and developmental spelling? *Reading Research Quarterly, 26*, 49 – 66.

Barnett, S. (1995). Long-term effects of early childhood programs on cognitive and school outcomes. *Future of Children, 5*, 25 – 50.

Bierman, K. L., Greenberg, M. T. & the Conduct Problems Prevention Research Group. (1996). Social skill training in the FAST Track program. In R. D. Peters & R. J. McMahon (Eds.), *Preventing childhood disorders, substance abuse, and delinquency* (pp. 65 – 89). Newbury Park, CA: Sage.

Blair, C. (2002). School readiness: Integrating cognition and emotion in a neurobiological conceptualization of child functioning at school entry. *American Psychologist, 57*, 111 – 127.

Blair, C. (2006). How similar are fluid cognition and general intelligence? A developmental neuroscience perspective on fluid cognition as an aspect of human cognitive ability. *Behavioral and Brain Sciences, 29*, 109 – 160.

Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read: A causal connection. *Nature, 301*, 419 – 421.

Brownell, R. (2000). *Expressive one-word picture vocabulary test manual*. Novato, CA: Academic Therapy Publications.

Bryant, P. E., MacLean, M., Bradley, L. L., & Crossland, J. (1990). Rhyme and alliteration, phoneme detection, and learning to read. *Developmental Psychology, 26*, 429 – 438.

Byrne, B., Fielding-Barnsley, R., & Ashley, L. (2000). Effects of preschool phoneme identity training after six years: Outcome level distinguished from rate of response. *Journal of Educational Psychology, 92*, 659 – 667.

Campbell, S. B., & von Stauffenberg, C. (2008). Child characteristics and family processes that predict behavioral readiness for school. In A. Crouter & A. Booth (Eds.), *Early disparities in school readiness: How families contribute to transitions into school* (pp. 225 – 258). Mahwah, NJ: Erlbaum.

Catts, H. W., Fey, M. E., Zhang, X., & Tomblin, J. B. (1999). Language basis of reading and reading disabilities: Evidence from a longitudinal investigation. *Scientific Studies of Reading, 3*, 331 – 361.

Conduct Problems Prevention Research Group. (1995). *Teacher social competence scale technical report*. Retrieved August 9, 2008, from the Fast Track Project Web site: http://www.fasttrackproject.org

Consortium on the School-Based Promotion of Social Competence. (1994). The school-based promotion of social competence: Theory, research, practice, and policy. In R. J. Haggerty, L. R. Sherrod, N. Garmezy, & M. Rutter (Eds.), *Stress, risk, and resilience in children and adolescents: Processes, mechanisms, and interventions* (pp. 268 – 316). New York: Cambridge University Press.

Coolahan, K., Fantuzzo, J., Mendez, J., & McDermott, P. (2000). Preschool peer interactions and readiness to learn: Relationships between classroom peer play and learning behaviors and conduct. *Journal of Educational Psychology, 92*, 458 – 465.

Crick, N. R., Casas, J. F., & Mosher, M. (1997). Relational and overt aggression in preschool. *Developmental Psychology, 33*, 579 – 588.

Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information processing mechanisms in children's social adjustment. *Psychological Bulletin, 115*, 74 – 101.

Denham, S., Bouril, B., & Belouad, F. (1994). Preschoolers' affect and cognition about challenging peer situations. *Child Study Journal, 24*, 1 – 21.

Denham, S. A., & Burton, R. (2003). *Social and emotional prevention and intervention programming for preschoolers*. New York: Kluwer-Plenum.

Dickinson, D. K., & Smith, M. W. (1994). Long-term effects of preschool teachers' book readings On low-income children's vocabulary and story comprehension. *Reading Research Quarterly, 29*, 105 – 1102.

Dodge, D. T., Colker, L. J., & Heroman, C. (2002). *The creative curriculum for preschool* (4th ed.). Washington, DC: Teaching Strategies.

Domitrovich, C. E., Cortes, R., & Greenberg, M. T. (2007). Improving young children's social and emotional competence: A randomized trial of the preschool PATHS curriculum. *Journal of Primary Prevention, 28*, 67 – 91.

Domitrovich, C. E., Greenberg, M. T., Cortes, R., & Kusche, C. (1999). *Manual for the preschool PATHS curriculum*. South Deerfield, MA: Channing-Bete.

Duncan, G. J., Brooks-Gunn, J., & Klebanov, P. K. (1994). Economic deprivation and early childhood development. *Child Development, 65*, 296 – 318.

DuPaul, G. (1991). Parent and teacher ratings of ADHD symptoms: Psychometric properties in a community-based sample. *Journal of Clinical Child Psychology, 20*, 245 – 253.

Eccles, J. S., & Harold (1996). Family involvement in children's and adolescents' schooling. In A. Booth & J. F. Dunn (Eds.), *Family-school links: How do they affect educational outcomes?* (pp. 3 – 34). Mahwah, NJ: Erlbaum.

Elias, M. J., Zins, J. E., Weissberg, R. P., Frey, K. S., Greenberg, M. T., Haynes, N. M., et al. (1997). *Promoting social and emotional learning: Guidelines for educators*. Alexandria, VA: Association for Supervision and Curriculum Development.

Greenberg, M. T., Kusche, C. A., & Speltz, M. (1991). Emotional regulation, self control, and psychopathology: The role of relationships in early childhood. In D. Cicchetti & S. L. Toth (Eds.), *Internalizing and externalizing expressions of dysfunction: Rochester symposium on developmental psychopathology* (Vol. 2, pp. 21 – 66). Hillsdale, NJ: Erlbaum.

Hatcher, P. J., Hulme, C., & Ellis, A. W. (1994). Ameliorating early reading failure by integrating the teaching of reading and phonological skills: The phonological linkage hypothesis. *Child Development, 65*, 41 – 57.

Hughes, J. N., & Kwok, O. (2006). Classroom engagement mediates the effect of teacher-student support on elementary students' peer acceptance: A prospective analysis. *Journal of School Psychology, 43*, 465 – 480.

Iutcovich, J., Fiene, R., Johnson, J., Koppel, R., & Langan, F. (1997). *Investing in our children's future: The path to quality child care through the Pennsylvania child care/early childhood development training system*. Erie, PA: Keystone University Research Corporation.

Kaiser, A. P., Hancock, T. B., Cai, Xinsheng, Foster, E. M., & Hester, P. P. (2000). Parent-reported behavioral problems and language delays in boys and girls enrolled in Head Start classrooms. *Behavioral Disorders, 26*, 26 – 41.

Kellam, S. G., Rebok, G. W., Ialongo, N. S., & Mayer, L. S. (1994). The course and malleability of aggressive behavior from early first grade into middle school: Results of a developmental epidemiology-based preventive trial. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 35*, 259 – 281.

Ladd, G. W., Buhs, E. S., & Seid, M. (2000). Children's initial sentiments about kindergarten: Is school liking an antecedent of early childhood classroom participation and achievement. *Merrill-Palmer Quarterly, 46*, 255 – 279.

Ladd, G., & Profilet, S. (1996). The Child Behavior Scale: A teacher report measure of young children's aggressive, withdrawn, and prosocial behaviors. *Developmental Psychology, 32*, 1008 – 1024.

Lengua, L. J., Honorado, E., & Bush, N. R. (2007). Contextual risk and parenting as predictors of effortful control and social competence in preschool children. *Journal of Applied Developmental Psychology, 28*, 40 – 55.

Li-Grining, C. P. (2007). Effortful control among low-income preschoolers in three cities: Stability, change, and individual differences. *Developmental Psychology, 43*, 208 – 221.

Lonigan, C. J. (2006). Development, assessment, and promotion of preliteracy skills. *Early Education and Development, 17*, 91 – 114.

Lonigan, C. J., Burgess, S. R., & Anthony, J. L. (2000). Development of emergent literacy and early reading skills in preschool children: Evidence from a latent-variable longitudinal study. *Developmental Psychology, 36*, 596 – 613.

Lonigan, C. J., Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (2007). *TOPEL: Test of Preschool Early Literacy*. Austin, TX: Pro-Ed.

Lynch, K. B., Geller, S. R., & Schmidt, M. G. (2004). Multi-year evaluation of the effectiveness of a resilience-based prevention program for young children. *Journal of Primary Prevention, 24*, 335 – 353.

McClelland, M. M., Acock, A. C., & Morrison, F. J. (2006). The impact of kindergarten learning-related skills on academic trajectories at the end of elementary school. *Early Childhood Research Quarterly, 21*, 471 – 490.

National Research Council. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

Newcomer, P. L., & Hammill, D. D. (1997). *The test of language development – Primary* (3rd ed.). Austin, TX: Pro-Ed.

Pianta, R. C., & Stuhlman, M. W. (2004). Teacher-child relationships and children's success in the first years of school. *School Psychology Review, 33*, 444 – 458.

Ramey, C. T., & Ramey, S. L. (2004). Early learning and school readiness: Can early intervention make a difference? *Merrill-Palmer Quarterly, 50*, 471 – 491.

Ribordy, S., Camras, L., Stafani, R., & Spacarelli, S. (1988). Vignettes for emotion recognition research and affective therapy with children. *Journal of Clinical Child Psychology, 17*, 322–325.

Riggs, N. R., Greenberg, M. T., Kusche, C. A., & Pentz, M. A. (2006). The mediational role of neurocognition in the behavioral outcomes of a social-emotional prevention program in elementary school students: Effects of the PATHS curriculum. *Prevention Science, 7*, 91–102.

Roid, G. H., & Miller, L. J. (1997). *Leiter International Performance Scale–Revised (Leiter–R) manual*. Wood Dale, IL: Stoelting.

Scarborough, H. (2001). Connecting early language and literacy to later reading (dis)abilities: Evidence, theory and practice. In S. Neuman & D. Dickinson (Eds.), *Handbook of emergent literacy research* (pp. 97–110). New York: Guildford.

Schultz, D., Izard, C. E., & Bear, G. (2004). Children's emotion processing: Relations to emotionality and aggression. *Development and Psychopathology, 16*, 371–387.

Schweinhart, L. J. (2004). *The High/Scope Perry Preschool Study through age 40: Summary, conclusions, and frequently asked questions*. Retrieved May 8, 2008, from http://www.highscope.org/Research/PerryProject/perrymain.htm

Schweinhart, L. J., Montie, J., Xiang, Z., Barnett, W. S., Belfield, C. R., & Nores, M. (2005). Lifetime effects: The High/Scope Perry Preschool Study through age 40. *Monographs of the High/Scope Educational Research Foundation, 14*. Ypsilanti, MI: High/Scope Press.

Senechal, M., & LeFevre, J. (2002). Parental involvement in the development of children's reading skill: A five-year longitudinal study. *Child Development, 73*, 445–460.

Shaywitz, B. A., & Shaywitz, S. E. (2005). Dyslexia (specific reading disability). *Biological Psychiatry, 57*, 1301–1309.

Shure, M. B., & Spivack, G. (1982). Interpersonal problem-solving in young children: A cognitive approach to prevention. *American Journal of Community Psychology, 10*, 341–356.

Smith-Donald, R., Raver, C. C., Hayes, T., & Richardson, B. (2007). Preliminary construct and concurrent validity of the preschool self-regulation assessment (PSRA) for field-based research. *Early Childhood Research Quarterly, 22*, 173–187.

Stanovich, K. E., & Siegel, L. S. (1994). The phenotypic performance profile of reading disabled children: A regression based test of the phonological core variables difference model. *Journal of Educational Psychology, 86*, 24–53.

Torgesen, J. K. (1998). Catch them before they fall: Identification and assessment to prevent reading failure in young children. *American Educator, 22*, 32–39.

U.S. Department of Health & Human Services (2005). *Head start impact study: First year findings*. Retrieved May 5, 2008, from http://www.acf.hhs.gov/programs/opre/hs/impact_study/reports/

Wasik, B. A., & Bond, M. A. (2001). Beyond the pages of a book: Interactive book reading and language development in preschool classrooms. *Journal of Educational Psychology, 93*, 243–250.

Wasik, B. A., Bond, M. A., & Hindman, A. (2006). The effects of a language and literacy intervention on Head Start children and teachers. *Journal of Educational Psychology, 98*, 63–74.

Webster-Stratton, C., Reid, J., & Hammond, M. (2004). Treating children with early-onset conduct problems: Intervention outcomes for parent, child, and teacher training. *Journal of Clinical Child and Adolescent Psychology, 33*, 105–124.

Weikart, D. P., & Schweinhart, L. J. (2005). The high/scope curriculum in early childhood care and education. In J. L. Roopnarine & J. E. Johnson (Eds.), *Approaches to early childhood education* (4th ed., pp. 277–294). Upper Saddle River, NJ: Prentice Hall.

Werthamer-Larsson, L., Kellam, S., & Wheeler, L. (1991). Effect of first-grade classroom environment on shy behavior, aggressive behavior, and concentration problems. *American Journal of Community Psychology, 19*, 585–602.

Whitehurst, G. J., Arnold, D., Epstein, J. N., Angell, A. L., Smith, M., & Fischel, J. E. (1994). A picture book reading intervention in daycare and home for children from low-income families. *Developmental Psychology, 30*, 679–689.

Whitehurst, G. J., Epstein, J. N., Angell, A. C., Payne, A. C., Crone, D. A., & Fischel, J. E. (1994). Outcomes of an emergent literacy intervention in Head Start. *Journal of Educational Psychology, 86*, 542–555.

Yoshikawa, H., & Knitzer, J. (1997). *Lessons from the field: Head Start mental health strategies to meet changing needs*. ERIC # ED423053.

Youngstrom, E., Wolpaw, J. M., Kogos, J. L., Schoff, K., Ackerman, B., & Izard, C. (2000). Interpersonal problem-solving in preschool and first-grade: Developmental changes and ecological validity. *Journal of Clinical Child Psychology, 4*, 589–602.

Child Development, November/December 2008, Volume 79, Number 6, Pages 1818–1832

Children's Cortisol and the Quality of Teacher–Child Relationships in Child Care

Jared A. Lisonbee
Washington State University

Jacquelyn Mize
Auburn University

Amie Lapp Payne
Washington State Department of Early Learning

Douglas A. Granger
Pennsylvania State University

Teacher–child relationships were examined as predictors of cortisol change in preschool children. Saliva for assays was collected from one hundred and ninety-one 4-year-olds (101 boys) in the mornings and afternoons on 2 days at child care, and before and after a series of challenging tasks and a teacher–child interaction session outside the classroom. Parents reported on children's temperament, teachers and children reported on teacher–child relationship quality, and observers rated group-level teacher insensitivity. Teacher-reported relationship conflict predicted cortisol increases during teacher–child interaction and teacher-reported overdependence predicted cortisol increases from morning to afternoon, even after controlling for individual teacher, child, and classroom characteristics. The findings extend earlier work by suggesting that cortisol change across the child-care day is influenced by teacher–child relationship characteristics.

A majority of preschool-aged children in the United States experience out-of-home care (Cappizzano, Adams, & Sonenstein, 2000). Researchers, policy makers, and parents have raised concerns about the potential negative effects of child care on children's health and development (National Institute of Child Health and Human Development [NICHD] Early Child Care Research Network, 2003). A series of studies suggests that many children in full-time out-of-home care experience atypical increases across the day in cortisol, the primary stress hormone in humans (Dettling, Gunnar, & Donzella, 1999; Tout, de Haan, Campbell, & Gunnar, 1998; Watamura, Donzella, Alwin, & Gunnar, 2003; Watamura, Sebanc, & Gunnar, 2002). Although the significance of this phenomenon is still being debated (Granger & Kivlighan, 2003), elevation of cortisol under some conditions has been linked to immune suppression (Boyce et al., 1995), problems with learning, memory (Blair, Granger, & Razza, 2005; Seeman & McEwen, 1996), negative emotionality, and problem behavior (Granger, Stansbury, & Henker, 1994; Smider et al., 2002).

An important research agenda is to identify the child and child-care characteristics that are associated with cortisol elevations. Individual differences in cortisol secretion reflect a confluence of forces including previous experience, child characteristics, contextual demands, and social relationships (Granger, Stansbury, et al., 1994; Stansbury & Gunnar, 1994). However, a consistent theme across studies is that relationships with caregivers are critical in helping young children regulate hypothalamic-pituitary-adrenal (HPA) activity (Gunnar, 1998; Gunnar & Donzella, 2002; Levine & Wiener, 1988). Secure attachment relationships appear to buffer infants' HPA reactivity to challenge (Gunnar, 1998; Nachmias, Gunnar, Mangelsdorf, Parritz, & Buss, 1996). Children in child care develop attachment-type relationships with their care providers (Howes, Galinsky, & Kontos, 1998), and these relationships may serve a similar protective function (Gunnar, 1998) for young children (Cole, Michel, & Teti, 1994). Positive teacher relationships may provide external coping resources and an environment of perceived safety, whereas problematic relationships may affect children's physiological functioning both by failing to buffer the potential

---

This article is based on work supported by the National Science Foundation under Grant 0126584 and the Agricultural Experiment Station (ALA010-008) to J.M. and G. Pettit, and by Graduate Dissertation Research Awards from the Auburn University Graduate School to J.A.L. and A.L.P. We thank the children, teachers, and parents involved in this study as well as the numerous undergraduate and graduate research assistants who assisted in data collection and preparation. We also are grateful for helpful comments of anonymous reviewers on previous versions of the manuscript.

Correspondence concerning this article should be addressed to Jared A. Lisonbee, Department of Human Development, Washington State University, 501 Johnson Tower, Pullman, WA 99164-4852. Electronic mail may be sent to jlisonbee@wsu.edu.

© 2008, No claim to original US Government works
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7906-0020

stress of the classroom and by causing anxiety for the child during teacher–child interactions.

The purpose of this study was to examine the role of teacher–child relationships in influencing cortisol levels in young children in child care. Cortisol changes were measured in two settings: (a) from morning to afternoon in the classroom and (b) outside the classroom across a series of challenging tasks and a teacher–child interaction. We expected that children with poorer quality relationships with teachers would show greater cortisol increases from morning to afternoon in the classroom and over the teacher–child interaction episode outside of the classroom. To assess cortisol changes across the day, saliva samples were collected from children enrolled in full-day community child care in the mornings and afternoons of 2 separate days. Because many factors operating in a child-care classroom may influence children's HPA axis functioning, we also collected saliva from children in an out-of-class activity consisting of two segments. The first segment involved a series of mildly challenging tasks selected to simulate events young children might face in the classroom setting. The second segment consisted of a nonchallenging interaction with the child's teacher. The teacher–child interaction session allowed for the isolation of teacher–child relationship effects from potential classroom influences. We also assessed potential child-, measurement-, and classroom-level influences on children's HPA activity. These included child age (Gunnar, Tout, de Haan, Pierce, & Stansbury, 1997), sex (Kirschbaum, Wust, & Hellhammer, 1992), temperament (Davis, Donzella, Krueger, & Gunnar, 1999; Gunnar, 1994), family socioeconomic status (SES; Lupien, King, Meaney, & McEwen, 2000, 2001), time of day (Tout et al., 1998), group size (Legendre, 2003), and teacher characteristics (Dettling, Parker, Lane, Sebanc, & Gunnar, 2000).

Method

Participants

Child participants were enrolled in a larger study examining associations between preschool experience and the transition to kindergarten. All recruitment and data collection procedures were approved by Auburn University's institutional review board. Twelve child-care centers in two adjacent small communities in the Southeastern United States participated. The centers served children from low-income families ($n = 2$), served employees of a regional hospital ($n = 1$), served employees of a public school system ($n = 1$), were operated by a church ($n = 1$), or were private/for profit ($n = 7$). Children in 4-year-old classrooms (i.e., children who would attend kindergarten the following year) were eligible to participate. Parents of 507 children enrolled at the centers (81% of possible participants) provided consent over a 3-year time span. Information provided by center administrators indicated comparable age, race, and SES characteristics for participants and nonparticipants. The ethnic composition of the sample was similar to that of the county according to the 2000 census (61% Caucasian, 33% African American, and 6% other ethnicity). Single-parent households were a minority (30%), and most families (98.7%) had at least one parent who was employed or a student. Among families with at least one employed adult, SES was measured using occupational prestige scores based on parents' reported employment (Entwisle & Astone, 1994). Scores ranged from 25 (e.g., unskilled labor, janitor) to 93 (professional), with a median of 68.

Because of the time and labor required to collect the physiological and relationship data, we selected a subgroup of approximately one third of the participants (five to six classrooms per year in each of 3 years) as a *focus sample* for the more labor-intensive portions of the study. Parent, teacher, and classroom-level data were collected from all the classrooms each year. Children in the focus classrooms also provided saliva samples, were interviewed about their relationships with their teachers, and participated in the individual challenge and teacher–child interaction sessions. This study uses only information collected from the focus sample.

In the 1st year of the study, focus classrooms were selected on the basis of informal judgments of quality to take advantage of the broadest possible range of child-care environments. In subsequent years, classrooms were selected to maximize the number of different classes and teachers involved across the 3 years of the study. Classrooms in 2 of the 12 centers never served as focus classrooms at the request of center administrators. Children attending the 2 centers not participating in the physiological and teacher–child assessment portions of the study differed from the rest of the sample only in maternal education, with mothers in the nonfocus centers reporting higher education levels than mothers in focus centers (*Ms* = 5.07 and 4.38, respectively, on a 6-point scale). Five classrooms were used twice across the 3 years of the study, but these classrooms had different teachers and children each year.

The resulting focus sample consisted of 191 children (90 girls) ranging in age from 43 to 67 months ($M = 53$, $SD = 4.1$ months). The focus sample was 66% Caucasian, 25% African American, and 9% mixed or other racial group. Focus and nonfocus samples differed by age and race, with focus sample children being older ($M = 53$ months vs. 52 months), $F(1, 506) = 10.57$, $p = .001$, and less likely to be African American (25% vs. 38%; $\chi^2 = 9.02$, $p < .05$).

Procedures

Saliva collection procedures. Children provided saliva samples for cortisol analysis in two settings. First, to assess morning and afternoon cortisol levels, and morning to afternoon cortisol changes, samples were collected in children's classrooms in winter (February) and spring (May). Additionally, children individually participated in an out-of-class assessment consisting of a series of mildly challenging tasks for 30 min (challenge task) followed by an interaction with their teacher for approximately 20 min (teacher–child interaction). The challenge task and teacher–child interaction sessions took place in the spring in a familiar quiet room in the center (e.g., a library or an unused classroom). Children provided samples immediately before (*pretask*) and following (*posttask*) the challenge task and again after the teacher–child interaction (*postinteraction*).

All saliva samples were collected at least 30 min after children had consumed any food. Twenty minutes prior to sample collections, children drank a cup of water (4–8 oz) to rinse the oral mucosa free of food residue and particulate matter and to increase hydration. Prior to the winter saliva collection, researchers taught children to provide unstimulated saliva by passive drool into 3-oz plastic cups. Immediately following each sample collection, researchers poured the saliva from the cups into 2-ml plastic cryogenic vials. Samples were sealed in plastic storage bags and placed on ice for transport to a storage freezer, where they remained at −20°C until shipped by overnight delivery on dry ice to Salimetrics Laboratories (State College, PA), where samples were stored at −80°C until assayed.

Morning, afternoon, and daily cortisol change. For classroom samples, children sat in small groups (four to six) at a table supervised by research assistants. Morning saliva collection occurred between 8:00 and 11:00 a.m., with 85% being collected immediately before children had morning snack between 8:30 and 9:30 a.m. ($M = $ 9:00 a.m.). There were no significant differences in a.m. or cortisol change scores for samples collected within this time frame and those collected slightly earlier or later. Following the logic of Shirtcliff, Granger, Booth, and Johnson (2005), winter and spring values were averaged to form morning and afternoon indexes of cortisol for primary analyses (*average daily cortisol change*). For children with full a.m. and p.m. cortisol data from only one collection day, the single a.m. and p.m. cortisol values were used, resulting in an n of 182 children. Unless otherwise specified, the average daily cortisol change was used in analyses.

Due to schedule variations and constraints at different centers, p.m. saliva collection times ranged from 1:00 to 4:00 p.m. and time between a.m. and p.m. samples varied widely (range = 3–7.25 hr, $M = 5.5$ hr, $SD = .75$ hr). Longer times between a.m. and p.m. sample collections were associated with greater cortisol increases across the day ($r = .16$, $p < .05$). Also because of scheduling constraints, 59% of afternoon saliva samples were collected within 30 min of children waking from nap. Consistent with previous findings (Watamura et al., 2002), cortisol levels were slightly lower in samples collected shortly after naptime (within 30 min, $M = .17$) than in samples collected later in the afternoon ($M = .20$), $F(1, 181) = 3.78$, $p < .10$. Nap duration (estimated from schedules provided by the centers) was associated with afternoon cortisol levels ($r = .21$, $p < .01$) and average daily cortisol change ($r = .22$, $p < .01$). Observer ratings of rest quality during naptime were collected in Year 3 of the study, but rest quality ratings were not significantly related to p.m. cortisol values or to daily change.

Challenge task. Children participated in a series of five developmentally salient tasks intended to provoke mild stress, frustration, or disappointment (Stansbury & Gunnar, 1994). Research assistants generated a randomized list of participating children from each classroom to determine the order of child participation. All children were eager to participate in the activity and no child declined. For all but 3 children, challenge task assessments occurred in the morning (M start time = 9:30 a.m., $SD = 43$ min). Children were accompanied to a room in the center by a familiar research assistant, seated at a small table, and immediately asked to provide a saliva sample (*prechallenge*). Children then participated in the challenge tasks, which consisted of a coordination game in which they were asked to drop plastic bears through a small hole in a canister, a disappointment experience in which they expected to receive an attractive toy as a prize but were given a broken toy instead (Cole, Zahn-Waxler, & Smith, 1994), an impossible puzzle task (Smiley & Dweck, 1994), a delay of gratification task (Kochanska & Aksan, 1995), and an inhibitory motor activity task (Kochanska, Murray, & Harlen, 2000). The total time to complete all tasks was approximately 30 min ($M = 31.8$ min, range = 24–43 min). Following the challenge tasks, children provided a second saliva sample (*postchallenge*). *Challenge change* cortisol was computed as postchallenge cortisol − prechallenge cortisol.

Teacher–child interaction session. Immediately after the postchallenge saliva collection, the child's teacher came into the room to interact with the child in two tasks intended to be nonstressful. First, teachers were asked to guide the child in duplicating a printed pattern with colored blocks, and the teacher and child were left alone for 5 min to complete the task. When the researcher returned to the room, he or she removed the blocks and pattern and gave the teacher a picture book with no printed text (Mayer & Mayer, 1975) to "read" with the child. After the teacher and child finished the book, the teacher returned to the classroom and the child provided a third saliva sample (postinteraction). On average, about 30 min elapsed between postchallenge and postinteraction saliva collections ($M = 29$ min, range = 19–40 min). After the final saliva collection, researchers escorted the child back to the classroom. *Interaction change* cortisol was computed as postinteraction cortisol − postchallenge cortisol.

Behavioral and Relationship Assessments

Teacher–child relationship quality. Teachers completed the Student–Teacher Relationship Scale (STRS; Pianta, Steinberg, & Rollins, 1995). The STRS contains 26 items assessing teachers' perceptions of relationships with individual children. Items are rated on 5-point scales. Following Pianta et al. (1995), three relationship scales were derived: *closeness* (11 items; $\alpha = .79$; e.g., "I share an affectionate, warm relationship with this child"), *conflict* (12 items; $\alpha = .90$; e.g., "This child and I always seem to be struggling with each other"), and overdependence or *clingy* (5 items; $\alpha = .73$; e.g., "This child asks for my help when he/she doesn't really need it"). Thirteen of 16 lead teachers completed questionnaires about their relationships with 155 participating children.

Child perceptions of relationship quality. Children's perceptions of their relationships with their lead teachers were assessed in individual interviews using the Feelings About My School and Teacher (FAMST) interview (Payne, 2003). Children participated in the FAMST interviews during the spring of the academic year. The FAMST is based on a modification of the Feelings About Myself and Peers (Meece, Colwell, & Mize, 2007) and the Berkeley Puppet interviews (Measelle, Ablow, Cowan, & Cowan, 1998). Representative items include, "Does [child's teacher] smile at you?" and "Do you like [teacher]?" For most questions, children answered verbally and by pointing to cards on which were mounted no, a few, or many milk bottle caps, corresponding to *not at all*, *some or a little bit*, and *a lot*, respectively. Ten items relating to the child's evaluation of his or her relationship with the teacher were aggregated to form a single scale with higher values (3s) reflecting more negative assessments (child-reported negativity, $\alpha = .69$). Pilot testing ($n = 19$) indicated adequate test–retest reliability (.81) over 2 weeks (Payne, 2003).

Child-level variables. Parents reported on temperament using the Children's Behavior Questionnaire (Rothbart, Ahadi, Hershey, & Fisher, 2001). Temperament dimensions examined in this study were shyness ($\alpha = .80$), fearfulness ($\alpha = .69$), and surgency (a composite of the high pleasure seeking, impulsive, high activity level, and [reversed] shyness scales; $\alpha = .70$). Parent evaluations of child temperament were available from 153 parents of focus sample children. Means for parent-reported temperament variables were similar to those reported in Rothbart et al. (2001). Additional child-level variables examined as potential influences on HPA activity included sex, age, and family SES.

Classroom-level variables. Group size was computed as number of children on class rolls plus the number of teachers typically present as reported by observers (Legendre, 2003). Teachers' insensitivity at the classroom level was assessed using observer reports on the Caregiver Interaction Scale (CIS; Arnett, 1989); the CIS consists of 26 items assessing teacher insensitivity (e.g., "threatens children to control them," and "speaks warmly to the children" [reverse scored]) rated on 4-point scales. Interrater reliability based on 40% of the classrooms in the study was adequate ($\rho_1 = .81$, $p < .001$) and internal consistency was good ($\alpha = .96$). Most teachers were rated as moderately sensitive at the classroom level ($M = 2.13$). However, teachers for 15% of the sample ($n = 3$ teachers) received average ratings greater than 3 on the scale on which 4 represented maximum insensitivity.

Measurement-level variables. On the 2 classroom saliva collection days, parents reported children's wake-up times as they dropped their child off at child care to control for diurnal cortisol patterns. Because some children carpooled or were bussed to the centers, wake-up times were available for only 90 of 164 children providing samples in the winter (55%) and 114 of 166 in the spring (69%). For children with missing wake-up data, we substituted an estimated wake-up time based on the average of wake-up times reported by parents on other saliva collection days and on questionnaires ($\alpha = .83$). There were no significant differences in any study variables between children with actual and substituted wake-up times. Wake-up times were typically not collected on the days when children participated in the out-of-class activities. To control for diurnal cortisol fluctuations for the out-of-class teacher–child interaction analysis, we used the average of available wake-up times reported by parents on saliva collection days and questionnaires.

Salivary Cortisol Determination

All samples were assayed for salivary cortisol using a highly sensitive enzyme immunoassay U.S. Food and Drug Administration (510k) cleared for use as an in vitro diagnostic measure of adrenal function (Salimetrics). The test used 25 μl of saliva, had a lower limit of sensitivity of 0.007 μg/dl, range of sensitivity from 0.007 to 1.8 μg/dl, and average intra- and interassay coefficients of variation of less than 5% and 10%. The average of duplicate assays for each sample was used in all analyses and units are reported in μg/dl (micrograms per deciliter). Cortisol values were positively skewed and subject to logarithmic transformation ($\log_{10}$). All analyses were performed using transformed cortisol values, but to ease interpretation, untransformed μg/dl units are reported in the text, tables, and figures.

Analytical Strategy

The data analyzed in this study represent multiple layers of information (i.e., repeated cortisol measurements from children sharing teacher and classroom influences). Given the nested nature of the data, we selected hierarchical multivariate linear modeling (HMLM; Raudenbush & Bryk, 2002; Raudenbush, Bryk, Cheong, & Congdon, 2004) to examine our primary research questions. HMLM uses multiple covariance matrices derived from different levels of data to accommodate repeated measurements within individuals who share group membership.

Results

Descriptive Analyses

Average daily cortisol change. Some children were absent or only attended half-day on days saliva was collected, so complete cortisol data were not available for all children (8%–10% missing at each collection). Data from three additional children were excluded because extreme values suggested interference or contamination. In winter and spring there were, respectively, 164 and 166 children with both morning and afternoon cortisol data. On average, children's cortisol levels did not change across the child-care days for either the winter or the spring assessments (see Table 1). As expected, however, there was considerable variation among children, with 51.2% and 45.8% showing cortisol increases greater than 1 *SD* of the mean across the day in winter and spring, respectively. Some children (29%) exhibited cortisol increases at both assessments. Winter and spring morning cortisol values were significantly associated ($r = .30$, $p < .001$), but afternoon values were only marginally related ($r = .13$, $p = .10$).

Challenge and interaction task cortisol. One hundred ninety-one children participated in the challenge task procedure. Five children were either unwilling (2) or unable (3) to give sufficient saliva volume for analysis. Four children had at least one extreme cortisol value (> 3 *SD* above the mean) excluded from the data. Descriptive information about challenge task cortisol values is given in Table 1. A repeated measures analysis of variance confirmed that cortisol levels declined, with postchallenge cortisol significantly lower than prechallenge, $F(1, 180) = 69.14$, $p < .001$, and postinteraction cortisol significantly lower than both pre- and postchallenge levels, $Fs(1, 180) = 69.65$ and 12.39, respectively, both $ps < .01$.

Relationship measures. In general, teachers described relationships with children as very close and relatively free of conflict, and children were seen as being moderately clingy in the relationships (see Table 1). Children's responses to the FAMST suggested that they saw their relationships with teachers as being of moderate quality. Teacher-rated closeness was negatively skewed and a quadratic transformation was used to normalize data. Child reports of negative relationships with teachers had a positive skew, and a $\log_{10}$ transformation was applied. Children and teachers agreed modestly about the quality of their relationships (see Table 2).

Hierarchical Analyses: Relationship Quality and Cortisol

Repeated measures hierarchical linear modeling (HLM 6; Raudenbush et al., 2004) was used to examine whether cortisol changes across the day and across the teacher–child interaction were associated with teacher–child relationship quality. Bivariate correlations revealed associations between cortisol and time since waking for the a.m., p.m., and prechallenge samples ($rs = .35$, .49, and .14, $ps < .001$, .001, and .10, respectively). Additionally, an association at the trend level was found for parent-reported fearful temperament and cortisol change across the challenge and teacher–child interaction tasks (see Table 3). Consequentially, time since

Table 1
Descriptive Statistics for Study Variables Prior to Transformation

	N	M (Mdn)	SD	Minimum to Maximum
Cortisol				
Winter a.m.[a]	171	0.17	0.12	0.06 to 0.93
Winter p.m.[a]	168	0.17	0.11	0.05 to 0.83
Winter daily change[a]	164	0.00	0.14	−0.64 to 0.44
Spring a.m.[a]	170	0.19	0.13	0.04 to 0.78
Spring p.m.[a]	171	0.19	0.11	0.04 to 0.65
Spring daily change[a]	166	−0.01	0.16	−0.62 to 0.57
Prechallenge[a]	181	0.15	0.09	0.04 to 0.59
Postchallenge[a]	182	0.12	0.08	0.03 to 0.57
Postinteraction[a]	182	0.11	0.05	0.03 to 0.42
Interaction change[a]	182	−0.02	0.05	−0.25 to 0.12
Teacher–child relationship quality				
Teacher-reported close	155	4.34	0.52	2.4 to 5.0
Teacher-reported clingy	155	2.46	0.92	1.0 to 4.8
Teacher-reported conflict	155	1.99	0.83	1.0 to 5.0
Child-reported negative	176	1.56	0.37	1.0 to 3.0
Control variables—child level				
Fearful temperament	153	4.03	1.2	1.0 to 6.67
Shy temperament	153	3.69	1.35	1.0 to 6.83
Surgent temperament	153	4.61	0.75	2.74 to 6.68
Child age in months	191	53	4.20	36 to 67
Family socioeconomic status	148	(65)		25 to 93[c]
Control variables—measurement				
Average wake to saliva a.m.[b]	173	2 hr 25 min	40 min	12 min to 4 hr 14 min
Average wake to saliva p.m.[b]	173	8 hr	51 min	5 hr 21 min to 10 hr 12 min
Duration of saliva collection	187	5 min	3 min	< 1 min to 18 min
Control variables—classroom level				
Group size	16	(17)		11 to 24
Teacher insensitive	16	(2.04)		1.31 to 3.38

[a]Cortisol values indicate mean of duplicate assay values in μg/dl prior to logarithmic transformation to correct skew. [b]Mean is hours and minutes between child wake up and saliva collection averaged over winter and spring. [c]Does not include families reporting that there were no employed adults in their households.

waking and fearfulness were controlled in the HLM models.

The HMLM analysis used multiple data covariance matrices to examine associations across measurement (Level 1), child (Level 2), and classroom (Level 3) influences. Level 1 data (measurement level) consisted of the cortisol levels at each assessment point as the outcome variable (each assessment occupied one row in the data set), along with the time since waking for each sample collected, and to assess the change (slope) across the interval of interest (e.g., a.m. to p.m.; postchallenge to postinteraction),.a dummy-coded indicator variable for the measurement of interest (e.g., the afternoon cortisol level for the average daily change analysis; see Raudenbush et al., 2004). The second level (child level) consisted of data describing the teacher–child relationship, sex, and parent-rated fearful temperament. Level 2 (child level) variables were modeled to predict both Level 1 intercept values and the slope (cortisol change) for the time period of interest (morning to afternoon or across the teacher–child interaction). Because different numbers of teachers and children reported on relationship quality, separate HMLM analyses were performed for the teacher and child relationship reports (*ns* = 155 and 176, respectively). Level 3 (classroom level) variables included classroom group size and teacher sensitivity.

Relationships and daily cortisol. Teacher reports of child clinginess were related to higher afternoon cortisol levels and average daily cortisol change. Results of the HMLM analyses confirmed that these associations (see Table 4) were independent of classroom, individual child, and other relationship measures. No other teacher- or child-rated relationship

Table 2

Correlations Among Teacher–Child Relationship Quality, Individual Child, and Classroom-Level Measures

Measures	2	3	4	5	6	7	8	9	10	11
Relationship quality										
1. Teacher-reported close	.18*	−.43***	−.20**	.18*	.01	.02	.11	−.02	−.08	−.05
2. Teacher-reported clingy		.18*	−.06	$-.15^{\dagger}$	.15	−.02	.05	−.01	−.05	.08
3. Teacher-reported conflict			.20**	−.21**	.06	−.10	−.11	.05	.06	−.07
4. Child-reported negativity				.12	−.06	$-.14^{\dagger}$	−.13	.06	−.01	.17*
Individual child										
5. Age					−.23**	−.12	.01	.11	.05	.03
6. Socioeconomic status						.23**	−.01	−.03	−.10	−.10
7. Fearful							.19*	−.22**	−.01	−.03
8. Shy								−.73***	−.01	.05
9. Surgent									.08	−.03
Classroom level										
10. Group size										−.33***
11. Teacher insensitivity										

$^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

variables were associated with average daily cortisol change.

Because of the potential influence of nap variables on afternoon cortisol levels, we tested a separate HMLM model with time since waking from nap in Level 2 and classroom nap duration in Level 3. Patterns of associations between relationship variables and average daily cortisol change were identical in the two models. For parsimony, we only report the results for the model without the nap variables.

Relationships and cortisol during challenge and teacher–child interaction. As seen in Table 3, children with more positive relationships with teachers (closer, less conflict, and less negative) began the challenge procedure with higher cortisol levels. Children with whom teachers reported more conflict and less closeness had greater cortisol *increases* across the teacher–child interaction (see Table 3). HMLM analyses (presented in Table 5) indicated that only teacher–child conflict predicted cortisol increases during the interaction session when classroom and child-level variables were included in the model.

Discussion

Within the past two decades, methodological advances in the noninvasive measurement of individual differences and intraindividual change in the psychobiology of stress have enabled developmental science to translate research with animals (Hofer, 2004; Meaney, 2005) into a basic understanding of early caregiving effects on children's physiology and behavior. Given the nature of the basic work with animals, it is not surprising that much of the research effort with children has focused on various aspects of mother–child relationship quality such as attachment (Ahnert, Gunnar, Lamb, & Barthel, 2004), conflict (Granger, Weisz, McCracken, Ikeda, & Douglas, 1996), and maternal behavior (Jones et al., 1997). Other research has examined differences between extreme groups defined by "natural experiments," such as child maltreatment and neglect (Cicchetti & Rogosch, 2001) and adoption and foster parenting (Dozier et al., 2006; Gunnar, Morison, Chisholm, & Schuder, 2001). This study extends this translational research by showing that relationships with care providers in out-of-home child-care settings during early childhood contribute to variability in children's HPA activity. Specifically, teacher reports of child overdependence predicted increases in cortisol levels across the child-care day, and reports of teacher–child relationship conflict predicted cortisol increases during a teacher–child interaction session.

Table 3
Correlations Among Cortisol and Relationship, Child, and Classroom Variables

Cortisol	Teacher–child relationship quality				Classroom level		Individual child				
	Teacher-reported close	Teacher-reported clingy	Teacher-reported conflict	Child-reported negativity	Groupsize	Teacher insensitive	Fearful	Shy	Surgent	Socioeconomic status	Age
Average classroom a.m.	.06	−.05	−.07	−.07	$-.14^{\dagger}$	$.14^{\dagger}$	.08	−.04	−.07	−.06	−.08
Average classroom p.m.	−.02	.22**	.02	−.06	.20**	.07	−.11	.00	−.07	−.08	−.05
Average daily cortisol change[a]	−.06	.20**	.08	−.02	.25**	−.08	−.13	.03	.02	.02	.01
Prechallenge	.17*	−.06	−.22**	−.18*	.03	−.04	.11	.03	−.05	−.02	−.05
Postchallenge	.08	−.03	−.13	−.12	.17*	−.06	.01	−.04	−.01	−.04	.04
Postinteraction	−.06	.02	.07	−.03	.24**	−.07	−.12	.00	.06	−.10	.05
Challenge change[a]	−.18*	.06	$.14^{\dagger}$	.09	$.13^{\dagger}$	−.05	$-.14^{\dagger}$	−.08	.07	−.13	.08
Interaction change[a]	$-.15^{\dagger}$	.03	.25**	$.13^{\dagger}$	.10	−.01	$-.16^{\dagger}$	.05	.06	−.01	.02

[a] Positive values indicate that cortisol increases over the time period.
$^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$, one-tailed.

Importantly, these associations persisted when controlling for child-, classroom-, and measurement-level factors known to influence variation in HPA activity.

Consistent with previous studies, approximately half of the children showed cortisol increases across the day at each of two assessments. However, descriptive analyses suggested considerable within-child variability in cortisol patterns, with only 29% of children experiencing cortisol increases across the day in both winter and spring. Morning cortisol had moderate intraindividual stability from winter to spring; afternoon values did not. These findings highlight the contextual sensitivity of the HPA axis (Flinn, 1999; Sapolsky, 2004). Shirtcliff et al. (2005) estimated that about 70% of variance in cortisol is attributable to situational influences. This study suggests that teacher–child relationship characteristics may be a contextual contributor to cortisol variability.

Teacher–Child Relationships as Contextual Influences

Different aspects of the teacher–child relationship were associated with HPA axis activation in different settings. Children rated by teachers as clingy or overdependent showed increases in cortisol over the day in the classroom. In contrast, children in more conflicted relationships showed elevated cortisol when interacting with the teacher outside of the classroom. In classrooms, children may be able to avoid interacting with teachers with whom they experience relationship difficulties. In the out-of-class activity, however, interaction was unavoidable. Previous research demonstrates that episodes of parent'–child conflict can be accompanied or followed by cortisol increases (Flinn, 1999; Granger et al., 1998), especially for children who perceive a lack of social control (Granger, Weisz, & Kauneckis, 1994). The animal literature suggests that subordinate, rather than dominant, members of dyads are more likely to experience HPA axis activation to conflict (Haller, Kiem, & Makara, 1996; Loving, Heffner, Kiecolt-Glaser, & Malarkey, 2004; Sapolsky, 2004). Certainly, children were the less powerful partners in the dyadic interaction sessions, and some children may have experienced these interactions as stressful, though they were not intended to be so.

The hypothesis that positive teacher–child relationships would predict decreases in cortisol across the day and during teacher–child interaction received modest support. Low conflict was related to greater *decreases* in cortisol across the teacher–child interaction, whereas teacher-reported closeness and child reports of less negativity in the relationship

Table 4

Summary of Hierarchical Multilevel Linear Modeling Repeated Measures Results of Teacher-Reported Relationship Qualities and Control Variables Predicting Average Daily Cortisol Change

	Cortisol intercept			Cortisol slope		
Variable	*B*	β	*t* value	*B*	β	*t* value
Level 1: Measurement						
Time since wake up	−.02	−.32	−1.75			
Level 2: Child						
Sex of the child	.00	.01	0.07	−.05	−.13	−1.32
Fearful temperament	.02	.08	1.14	−.03	−.15	−1.56
Teacher-reported conflict	.01	.02	−0.36	−.01	−.07	−0.54
Teacher-reported close	.00	.08	0.96	−.01	−.16	−1.45
Teacher-reported clingy	−.02	−.06	−0.88	.07	.31	3.00**
Level 3: Classroom						
Group size	.00	.06	0.67			
Teacher insensitive	.04	.15	1.50			

$^{**}p < .01$.

were marginally related to declines in cortisol during teacher–child interaction in bivariate analyses. When controlling for other child and classroom variables in the HMLM models, only teacher–child conflict predicted cortisol change during the teacher–child interaction. The modest links between teacher-reported closeness and children's cortisol levels may be explained by low variability in closeness ratings; teachers may be constrained by social desirability to describe relationships with children under their care in positive terms. Alternative methods for quantifying positive aspects of teacher–child relationships, such as observation of classroom interactions (Howes et al., 1998), should be explored.

Alternative Hypotheses

Temperament. It is possible that teacher-reported clinginess reflects assessments of behavioral temperament rather than relationship characteristics. Children described as clingy by teachers may lack confidence in their abilities to cope with challenges in the classroom. Similarly, children with whom teachers said they had more conflict may have been

Table 5

Summary of Hierarchical Multilevel Linear Modeling Repeated Measures Results of Teacher-Reported Relationship Qualities and Control Variables Predicting Cortisol Changes Across the Teacher–Child Interaction

	Cortisol intercept			Cortisol slope		
Variable	*B*	β	*t* value	*B*	β	*t* value
Level 1: Measurement						
Time since wake up	.02	.06	0.94			
Level 2: Child						
Sex of the child	.03	.07	0.84	−.04	−.09	−1.72
Fearful temperament	.00	.00	0.06	−.01	−.05	−1.05
Teacher-reported conflict	−.03	−.11	−1.22	.03	.20	2.19*
Teacher-reported close	.00	.06	0.71	−.00	−.02	−0.38
Teacher-reported clingy	−.00	−.01	−0.15	.00	.01	0.14
Level 3: Classroom						
Group size	.01	.22	3.12*			
Teacher insensitive	.00	.02	0.22			

$^{*}p < .05$.

the more difficult or active children who felt constrained in the interaction session. This interpretation is consistent with previous research linking temperament to children's cortisol at child care (Dettling et al., 1999; Gunnar, Sebanc, Tout, Donzella, & van Dulmen, 2003; Watamura et al., 2002).

By and large, however, the pattern of associations among study variables does not support the temperament interpretation. Had teachers based their relationship ratings on children's temperament, some correlations between relationship quality and parent temperament ratings would be expected. However, there was no significant correlation between any teacher-reported relationship scale and any parent-reported temperament dimension. In contrast, children and teachers agreed modestly about the quality of their relationships, and children rated relationships with less sensitive teachers as being more negative. This pattern suggests that both teachers' and children's reports captured some specifically dyadic aspects of their relationships.

Of the three temperament qualities examined in this study, only fearfulness was associated with any cortisol measure: More fearful children showed a trend toward greater decreases in cortisol over both the challenge task and the teacher – child interaction session. The nonsignificant associations between child characteristics and HPA activity are consistent with the notion that context is a stronger influence on cortisol than are individuals' traits (Shirtcliff et al., 2005), but contrary to some previous research reporting that children who are shy (Gunnar et al., 1997), have poorer effortful control, or have greater negative emotionality (Dettling et al., 1999; Dettling et al., 2000; Gunnar et al., 2003) are more likely to experience cortisol increases over the day at child care. This discrepancy may be attributed to measurement and sample characteristics. Some previous studies have used teacher reports of temperament (Gunnar et al., 2003; Watamura et al., 2002), or composites of teacher and parent reports (Dettling et al., 2000), and teacher reports may be more likely than parent reports (Manz, Fantuzzo, & McDermott, 1999) to correspond to children's behavioral and emotional characteristics at school (Culp, Howell, Culp, & Blankemeyer, 2001). However, parent reports of low effortful control and impulsivity also have been linked to cortisol increases across the day in children attending a university child-care program (Dettling et al., 1999). Parents in the current sample represented a wide range of economic and educational backgrounds and may have interpreted the temperament questionnaire differently than did parents in previous studies. In particular, the moderate internal reliability for the temperament scales in this study may have limited the ability to detect associations with children's cortisol.

Age. Younger preschool children are more likely to experience increases in cortisol over the day in child-care settings (Dettling et al., 1999), especially younger children in mixed-age groups (Legendre, 2003). Age was not associated with any cortisol measures in this study, perhaps because of the restricted age range of children. However, age may indirectly influence cortisol through children's experiences with their teachers. Teachers rated relationships with older children as being closer and less conflicted and older children were marginally less clingy than younger children. Examining age differences in social behavior may inform our understanding of children's physiological experience in classrooms.

Methodological influences relating to the teacher – child interaction. The fact that the challenge task always preceded the teacher – child interaction session is both a strength and a weakness of the current study. On the one hand, carryover effects of stress children may have experienced during the challenge task could influence cortisol during the teacher – child interaction. On the other hand, all children were exposed to a comparable experience prior to interacting with their teachers. This consistency allows for better isolation of the effects of the teacher – child relationship on cortisol than would be possible if children had been in their classrooms prior to the interaction.

Another methodological consideration is the timing of the out-of-class assessment. In this study, the teacher – child interaction occurred in the morning. Because of the typical decrease in cortisol levels across the day, the HPA axis may be more sensitive to stressors in the afternoon than in the morning (Dickerson & Kemeny, 2004). Higher morning cortisol may obscure stressor effects (Ramsay & Lewis, 2003) in part because of stronger genetic effects on morning as compared to afternoon cortisol levels (Bartels, de Geus, Kirschbaum, Sluyter, & Boomsma, 2003). Results may have been different had the challenge and interaction procedures occurred in the afternoon when, theoretically, the influence of context would have been more powerful. However, because of the timing of naps, snacks, variation in parent pickup time, and child-care staff shift changes, staging the teacher – child interaction in the afternoon was logistically impossible.

Classroom influences. Of all study variables, group size had the most consistent pattern of bivariate associations with cortisol levels and changes. Children in classrooms with more members had marginally lower morning cortisol but significantly higher afternoon cortisol and greater cortisol increases across the day. The afternoon and morning to afternoon cortisol change data are consistent with previous research showing higher cortisol in toddlers in larger groups (Legendre, 2003) and with studies of the effects of crowding on stress and HPA axis functioning (Nelson, 2005; Thiessen & Rogers, 1961). Larger groups are louder, offer more opportunities for conflict and competition for scarce resources (including teachers' time), and are more stressful for teachers (French, 1993). Teachers with larger classrooms were rated by observers as less sensitive to children's needs. Thus, class size may operate directly and indirectly (through teacher behavior) to influence children's cortisol.

More unexpected were associations between group size and higher cortisol in the out-of-class activity. Because each child had been out of the classroom for about half an hour prior to beginning the teacher interaction session, direct effects of classroom factors on cortisol during teacher–child interaction were expected to be minimal. Even so, children in larger groups had greater increases in cortisol over the challenge task and higher cortisol levels after the challenge and teacher interaction procedures. Stressful classroom events prior to the challenge procedure, which may have occurred more often in larger groups, may have continued to influence children's HPA axis activity even after children had been out of the classroom for half an hour or more. This seems an unlikely explanation, however, given that pre-challenge cortisol was not associated with group size. Perhaps, children in smaller groups develop better coping skills, more confidence, or are less reactive to stressful situations, all of which would support better adaptation to unexpected challenges.

Why children in larger groups had lower morning cortisol levels is not obvious, and, given the modest magnitude, it may be spurious. In contrast, some studies have documented low morning cortisol for children in center-based child care (Dettling et al.; Dettling et al., 2000), which affords some degree of confidence in the finding. The explanation may lie in family conditions that covary with child-care arrangements (NICHD Early Child Care Research Network, 1997). Perhaps parents who reported to work earlier tended to place their children in child-care arrangements serving larger groups so that children in larger groups were awake longer prior to the morning saliva collections. As expected, children in our study who had been awake longer had lower cortisol levels at the morning assessments. However, controlling for the amount of time children had been awake at the morning collections did not affect the magnitude of the association between group size and any cortisol value. The link between group size and lower morning cortisol could be a concern if it suggests disruption of the normal diurnal pattern (Gunnar & Vasquez, 2001).

Conclusions and Implications

This study used a community-based sample to replicate findings that a substantial proportion of preschool-aged children in full-day child care experience an afternoon cortisol increase, rather than decrease, across the day. We found that aspects of the teacher–child relationship predicted cortisol increases both in the classroom and during a dyadic teacher–child interaction outside the classroom, even after controlling for measurement-, child-, and classroom-level factors previously found to influence cortisol. Previous research demonstrates the importance of teacher–child relationships for young children's school and behavioral adjustment concurrently (Birch & Ladd, 1997) and longitudinally to early adolescence (Hamre & Pianta, 2001). This study demonstrated associations between teacher–child relationships and children's physiological functioning.

Whether elevated cortisol functions as more than simply a marker of children's immediate discomfort, anxiety, or exertion is not clear. Cortisol elevations in many circumstances are adaptive (Sapolsky, 2004) and a marker of social competence, particularly in new situations (Gunnar et al., 1997). Higher baseline cortisol levels among preschool-aged children predict internalizing symptoms in elementary school, even after controlling for early behavioral evidence of anxious–fearful behavior (Smider et al., 2002). Conversely, low baseline cortisol is a robust psychophysiological correlate of antisocial behavior (McBurnett, Lahey, Rathouz, & Loeber, 2000; Scarpa & Raine, 2006). Low morning cortisol appears to be a marker of current or previous adversity and is characteristic of adults with a history of childhood trauma (Weissbecker, Floyd, Dedert, Salmon, & Sephton, 2006) and children growing up in Eastern European orphanages (see Gunnar & Vazquez, 2001, for a review). Gunnar and Vasquez (2001) have argued that low morning cortisol, or hypocortisolism, in young children can reflect a downregulation of the HPA axis in response to chronic activation and appears to be a risk factor for problems in several domains, including attentional, behavioral, and health-related conditions. In contrast, animal studies suggest moderate levels of cortisol are necessary for supporting neural development (de Kloet, Oitzl, & Joëls, 1999). Identifying optimal cortisol levels in children and factors that contribute to healthy HPA axis functioning constitutes an important agenda for developmental research.

The modest across-time stability of cortisol measures in this study and the high moment-to-moment variability in cortisol in general (Shirtcliff et al., 2005) likely reflect unmeasured and transient experiences. In the child-care setting, these experiences would include peer conflict, injuries, and discipline encounters. The apparent importance of such events suggests the need for more fine-grained, moment-to-moment analysis of children's experiences in child care in conjunction with concurrent collection of multiple saliva samples to chart experience–cortisol links (Coste, Strauch, Letrait, & Bertagna, 1994; Flinn, 1999). This approach would be consistent with the view that more global aspects of child-care environments, such as group size and teacher sensitivity, are best conceptualized in terms of how they influence children's actual experiences (Pianta, 1999). Cortisol fluctuations in child care most likely reflect interplay between individual child, teacher, peer, and classroom influences.

Some have argued that concurrent examination of multiple physiological systems is necessary to fully understand the significance of these systems and their disrupted functioning for children's development (Bauer, Quas, & Boyce, 2002; Granger & Kivlighan, 2003). In particular, asynchrony in reactivity of the sympathetic-adrenal-medullary (SAM) and the HPA systems may mark or condition risk for adjustment problems (Bauer et al., 2002). The development of procedures to measure alpha-amylase, a marker of SAM activation (Chatterton, Vogelsong, Lu, Ellman, & Hudgens, 1996), in saliva makes it feasible to noninvasively asses both HPA and SAM activity in young children (Granger et al., 2006). Future research should incorporate other aspects of children's physiological functioning to better understand how child care influences children.

This study sheds additional light on an as yet incompletely understood phenomenon—disruption of the typical pattern of HPA activity among many young children attending full-day child care. Additionally, the data begin to situate child care–cortisol research in the context of a broader conceptual and empirical literature on the role of relationships in shaping baseline functioning and reactivity of physiological systems in developing organisms (Sanchez, Ladd, & Plotsky, 2001).

References

Ahnert, L., Gunnar, M. R., Lamb, M. E., & Barthel, M. (2004). Transition to child care: Associations with infant-mother attachment, infant negative emotion, and cortisol elevations. *Child Development*, *75*, 639–650.

Arnett, J. (1989). Caregivers in day-care centers: Does training matter? *Journal of Applied Developmental Psychology*, *10*, 541–552.

Bartels, M., de Geus, E. J. C., Kirschbaum, C., Sluyter, F., & Boomsma, D. I. (2003). Heritability of daytime cortisol levels in children. *Behavior Genetics*, *33*, 421–433.

Bauer, A. M., Quas, J. A., & Boyce, W. T. (2002). Associations between physiological reactivity and children's behavior: Advantages of a multisystem approach. *Journal of Developmental and Behavioral Pediatrics*, *23*, 102–113.

Birch, S. H., & Ladd, G. W. (1997). The teacher-child relationship and children's early school adjustment. *Journal of School Psychology*, *35*, 61–79.

Blair, C., Granger, D., & Razza, R. P. (2005). Cortisol reactivity is positively related to executive function in preschool children attending Head Start. *Child Development*, *76*, 554–567.

Boyce, W. T., Adams, S., Tschann, J. M., Cohen, F., Wara, D., & Gunnar, M. R. (1995). Adrenocortical and behavioral predictors of immune responses to starting school. *Pediatric Research*, *38*, 1009–1017.

Cappizzano, J., Adams, G., & Sonenstein, F. (2000). Child care arrangements for children under five: Variation across states. In *Assessing the new federalism: An Urban Institute program to assess changing social policies* (No. B-7). Washington, DC: The Urban Institute.

Chatterton, R. T., Vogelsong, K. M., Lu, Y., Ellman, A. B., & Hudgens, G. A. (1996). Salivary α-amylase as a measure of endogenous adrenergic activity. *Clinical Psychology*, *16*, 433–448.

Cicchetti, D., & Rogosch, F. A. (2001). The impact of child maltreatment and psychopathology on neuroendocrine functioning. *Development and Psychopathology*, *13*, 783–804.

Cole, P. M., Michel, M. K., & Teti, L. O. (1994). The development of emotion regulation and dysregulation: A clinical perspective. *Monographs of the Society for Research in Child Development*, *59*(Serial No. 240), 73–100.

Cole, P. M., Zahn-Waxler, C., & Smith, K. D. (1994). Expressive control during a disappointment: Variations related to preschoolers' behavior problems. *Developmental Psychology*, *30*, 835–846.

Coste, J., Strauch, G., Letrait, M., & Bertagna, X. (1994). Reliability of hormonal levels for assessing the hypothalamic-pituitary-adrenocortical system in clinical pharmacology. *British Journal of Clinical Pharmacology*, *38*, 474–479.

Culp, R. E., Howell, C. S., Culp, A. M., & Blankemeyer, M. (2001). Maltreated children's emotional and behavioral problems: Do teachers and parents see the same things? *Journal of Child and Family Studies*, *10*, 39–50.

Davis, E. P., Donzella, B., Krueger, W. K., & Gunnar, M. R. (1999). The start of a new year: Individual differences in salivary cortisol response in relation to child temperament. *Developmental Psychobiology*, *35*, 118–196.

de Kloet, E. R., Oitzl, M. S., & Joëls, M. (1999). Stress and cognition: Are corticosteroids good or bad guys. *Trends in Neuroscience*, *22*, 422–426.

Dettling, A. C., Gunnar, M. R., & Donzella, B. (1999). Cortisol levels of young children in full-day childcare centers: Relations with age and temperament. *Psychoneuroendocrinology*, *24*, 519–536.

Dettling, A. C., Parker, S. W., Lane, S., Sebanc, A., & Gunnar, M. R. (2000). Quality of care and temperament determine changes in cortisol concentrations over the day for young children in childcare. *Psychoneuroendocrinology*, *25*, 819–836.

Dickerson, S. S., & Kemeny, M. E. (2004). Acute stressors and cortisol responses: A theoretical integration and synthesis of laboratory research. *Psychological Bulletin*, *130*, 355–391.

Dozier, M., Manni, M., Gordon, M. K., Peloso, E., Gunnar, M. R., Stovall-McClough, K. C., et al. (2006). Foster children's diurnal production of cortisol: An exploratory study. *Child Maltreatment*, *11*, 189–197.

Entwisle, D. R., & Astone, N. M. (1994). Some practical guidelines for measuring youth's race/ethnicity and socioeconomic status. *Child Development*, *65*, 1521–1540.

Flinn, M. V. (1999). Family environment, stress, and health during childhood. In C. Panter-Brick & C. M. Worthman (Eds.), *Hormones, health, and behavior* (pp. 105–138). Cambridge, England: Cambridge University Press.

French, N. K. (1993). Elementary teacher stress and class size. *Journal of Research and Development in Education*, *26*, 66–73.

Granger, D. A., & Kivlighan, K. T. (2003). Integrating biological, behavioral, and social levels of analysis in early child development: Progress, problems, prospects. *Child Development*, *74*, 1058–1063.

Granger, D. A., Kivlighan, K. T., Blair, C., El-Sheikh, M., Mize, J., Lisonbee, J. A., et al. (2006). Integrating the measurement of salivary a-amylase into studies of child health, development, and social relationships. *Journal of Personal and Social Relationships*, *23*, 267–290.

Granger, D. A., Serbin, L. A., Schwartzman, A., Lehoux, P., Cooperman, J., & Ikeda, S. (1998). Children's salivary cortisol, internalizing behavior problems, and family environment: Results from the Concordia Longitudinal Risk Project. *International Journal of Behavioral Development*, 22, 707–728.

Granger, D. A., Stansbury, K., & Henker, B. (1994). Preschoolers' behavioral and neuroendocrine responses to social challenge. *Merrill-Palmer Quarterly*, *40*, 190–211.

Granger, D. A., Weisz, J. R., & Kauneckis, D. (1994). Neuroendocrine reactivity, internalizing behavior problems, and control-related cognitions in clinic-referred children and adolescents. *Journal of Abnormal Child Psychology*, *103*, 267–276.

Granger, D. A., Weisz, J. R., McCracken, J. T., Ikeda, S. C., & Douglas, P. (1996). Reciprocal influences among adrenocortical activation, psychosocial processes, and behavioral adjustment of clinic-referred children. *Child Development*, *67*, 3250–3262.

Gunnar, M. R. (1994). Psychoendocrine studies of temperament and stress in early childhood: Expanding current models. In J. E. Bates & T. D. Wachs (Eds.), *Temperament: Individual differences at the interface of biology and behavior* (pp. 175–198). Washington, DC: American Psychological Association.

Gunnar, M. R. (1998). Quality of early care and buffering of neuroendocrine stress reactions: Potential effects on the developing brain. *Preventive Medicine: An International Journal Devoted to Practice and Theory*, *27*, 208–211.

Gunnar, M. R., & Donzella, B. (2002). Social regulation of the cortisol levels in early human development. *Psychoneuroendocrinology*, *27*, 199–220.

Gunnar, M. R., Morison, S. J., Chisholm, K., & Schuder, M. (2001). Salivary cortisol levels in children adopted from Romanian orphanages. *Development and Psychopathology*, *13*, 611–628.

Gunnar, M. R., Sebanc, A. M., Tout, K., Donzella, B., & van Dulmen, M. M. H. (2003). Peer rejection, temperament, and cortisol activity in preschoolers. *Developmental Psychobiology*, *43*, 346–358.

Gunnar, M. R., Tout, K., de Haan, M., Pierce, S., & Stansbury, K. (1997). Temperament, social competence, and adrenocortical activity in preschoolers. *Developmental Psychobiology*, *31*, 65–85.

Gunnar, M. R., & Vazquez, D. M. (2001). Low cortisol and flattening of expected daytime rhythm: Potential indices of risk in human development. *Development and Psychopathology*, *13*, 515–538.

Haller, J., Kiem, D. T., & Makara, G. B. (1996). The physiology of social conflict in rats: What is particularly stressful? *Behavioral Neuroscience*, *110*, 353–359.

Hamre, B. K., & Pianta, R. C. (2001). Early teacher-child relationships and the trajectory of children's school outcomes through eighth grade. *Child Development*, *72*, 625–638.

Hofer, M. (2004). Developmental psychobiology of early attachment. In B. J. Casey (Ed.), *Developmental psychobiology* (pp. 1–28). Washington, DC: American Psychiatric Publishing.

Howes, C., Galinsky, E., & Kontos S. (1998). Child care caregiver sensitivity and attachment. *Social Development*, *7*, 25–36.

Jones, N. A., Field, T., Fox, N. A., Davalos, M., Mulphurs, J., Carraway, K, et al. (1997). Infants of intrusive and withdrawn mothers. *Infant Behavior and Development*, *20*, 175–186.

Kirschbaum, C., Wust, S., & Hellhammer, D. (1992). Consistent sex differences in cortisol responses to psychological stress. *Psychosomatic Medicine*, *54*, 648–657.

Kochanska, G., & Aksan, N. (1995). Mother-child mutually positive affect, the quality of child compliance requests and prohibitions, and maternal control as correlates of early internalization. *Child Development*, *66*, 236–254.

Kochanska, G., Murray, K. T., & Harlen, E. T. (2000). Effortful control in early childhood: Continuity and change, antecedents, and implications for social development. *Developmental Psychology, 36*, 220–232.

Legendre, A. (2003). Environmental features influencing toddlers' bioemotional reactions in day care centers. *Environment and Behavior, 35*, 523–549.

Levine, S., & Wiener, S. G. (1988). Psychoendocrine aspects of mother-infant relationships in nonhuman primates. *Psychoneuroendocrinology, 13*, 143–154.

Loving, T. J., Heffner, K. L., Kiecolt-Glaser, J. K., Glaser, R., & Malarkey, W. B. (2004). Stress hormone changes and marital conflict: Spouses' relative power makes a difference. *Journal of Marriage and the Family, 66*, 595–612.

Lupien, S. J., King, S., Meaney, M. J., & McEwen, B. S. (2000). Child's stress hormone levels correlate with mother's socioeconomic status and depressive state. *Biological Psychiatry, 48*, 976–980.

Lupien, S. J., King, S., Meaney, M. J., & McEwen, B. S. (2001). Can poverty get under your skin? Basal cortisol levels and cognitive function in children from low and high socioeconomic status. *Development and Psychopathology, 13*, 653–676.

Manz, P. H., Fantuzzo, J. W., & McDermott, P. A. (1999). The parent version of the preschool Social Skills Rating Scale: An analysis of its use with low-income, ethnic minority children. *School Psychology Review, 28*, 493–504.

Mayer, M., & Mayer, M. (1975). *One frog too many.* New York: Dial Press.

McBurnett, K., Lahey, B. B., Rathouz, P., & Loeber, R. (2000). Low salivary cortisol and persistent aggression in boys referred for disruptive behaviors. *Archives of General Psychiatry, 57*, 38–43.

Meaney, M. J. (2005). Development of animal models in neuroscience and molecular biology. In J. M. Oldham, A. E. Skodol, & D. S. Bender (Eds.), *The American Psychiatric Publishing textbook of personality disorders* (pp. 653–668). Washington, DC: American Psychiatric Publishing.

Measelle, J., Ablow, J., Cowan, P., & Cowan, C. (1998). Assessing young children's views of their academic, social, and emotional lives: An evaluation of the self-perception scales of the Berkeley Puppet Interview. *Child Development, 69*, 1556–1576.

Meece, D. W., Colwell, M. J., & Mize, J. (2007). Maternal emotion framing and children's social behavior: The role of children's feelings and beliefs about peers. *Early Child Development and Care, 177*, 295–309.

Nachmias, M., Gunnar, M. R., Mangelsdorf, S., Parritz, R. H., & Buss, K. (1996). Behavioral inhibition and stress reactivity: The moderating role of attachment security. *Child Development, 67*, 508–522.

Nelson, R. J. (2005). *An introduction to behavioral endocrinology.* Sunderland, MA: Sinauer.

National Institute of Child Health and Human Development Early Child Care Research Network. (1997). Familial factors associated with the characteristics of nonmaternal care for infants. *Journal of Marriage and the Family, 59*, 389–408.

National Institute of Child Health and Human Development Early Child Care Research Network. (2003). Does amount of time spent in child care predict socioemotional adjustment during the transition to kindergarten? *Child Development, 74*, 976–1005.

Payne, A. L. (2003). Measuring preschool children's perceptions of teacher-child relationship quality: Developing and testing the "Feelings About My School and Teachers" instrument. *Dissertation Abstracts International, 64* (2-A), 391.

Pianta, R. C. (1999). How the parts affect the whole: Systems theory in classroom relationships. In R. C. Pianta (Ed.), *Enhancing relationships between teachers and children* (pp. 23–43). Washington, DC: American Psychological Association.

Pianta, R. C., Steinberg, M., & Rollins, K. (1995). The first two years of school: Teacher-child relationships and deflections in children's classroom adjustment. *Development and Psychopathology, 7*, 295–312.

Ramsay, D., & Lewis, M. (2003). Reactivity and regulation in cortisol and behavioral responses to stress. *Child Development, 74*, 456–464.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods.* Newbury Park, CA: Sage.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., & Congdon, R. T., Jr. (2004). *HLM 6: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International.

Rothbart, M. K., Ahadi, S. A., Hershey, K. L., & Fisher, P. (2001). Investigation of temperament at three to seven years: The Children's Behavior Questionnaire. *Child Development, 72*, 1394–1408.

Sanchez, M. M., Ladd, C. O., & Plotsky, P. M. (2001). Early adverse experience as a developmental risk factor for later psychopathology: Evidence from rodent and primate models. *Development and Psychopathology, 13*, 419–449.

Sapolsky, R. M. (2004). *Why zebras don't get ulcers.* New York: Henry Holt.

Scarpa, A., & Raine, A. (2006). The psychophysiology of human antisocial behavior. In R. J. Nelson (Ed.), *Biology of aggression* (pp. 447–461). New York: Oxford.

Seeman, T., & McEwen, B. (1996). Impacts of social environment characteristics on neuroendocrine regulation. *Psychosomatic Medicine, 58*, 459–471.

Shirtcliff, E. A., Granger, D. A., Booth, A., & Johnson, D. (2005). Low salivary cortisol and externalizing problem behavior in youth. *Development and Psychopathology, 17*, 167–184.

Smider, N. A., Essex, M. J., Kalin, N. H., Buss, K. A., Klein, M. H., Davidson, R. J, et al. (2002). Salivary cortisol as a predictor of socioemotional adjustment during kindergarten: A prospective study. *Child Development, 73*, 75–92.

Smiley, P. A., & Dweck, C. S. (1994). Individual differences in achievement goals among young children. *Child Development, 65*, 1723–1743.

Stansbury, K., & Gunnar, M. R. (1994). Adrenocortical activity and emotion regulation. *Monographs of the Society for Research in Child Development*, 59(2/3, Serial No. 240), 108–134.

Thiessen, D. D., & Rodgers, D. A. (1961). Population density and endocrine function. *Psychological Bulletin*, *58*, 441–451.

Tout, K., de Haan, M., Campbell, E. K., & Gunnar, M. R. (1998). Social behavior correlates of cortisol activity in child care: Gender differences and time of day effects. *Child Development*, *69*, 1247–1262.

Watamura, S. E., Donzella, B., Alwin, J., & Gunnar, M. R. (2003). Morning to afternoon increases in cortisol concentrations for infants and toddlers at child care: Age differences and behavioral correlates. *Child Development*, *74*, 1006–1020.

Watamura, S. E., Sebanc, A. M., & Gunnar, M. R. (2002). Rising cortisol in child care: Relations with nap, rest, and temperament. *Developmental Psychobiology*, *40*, 33–42.

Weissbecker, I., Floyd, A., Dedert, E., Salmon, P., & Sephton, S. (2006). Childhood trauma and diurnal cortisol disruptions in fibromyalgia syndrome. *Psychoneuroendocrinology*, *31*, 312–324.

Child Development, November/December 2008, Volume 79, Number 6, Pages 1833–1852

Longitudinal Effects of Student Mobility on Three Dimensions of Elementary School Engagement

Diana H. Gruman
Western Washington University

Tracy W. Harachi, Robert D. Abbott, Richard F. Catalano, and Charles B. Fleming
University of Washington

Working within the developmental science research framework, this study sought to capture a dynamic and complex view of student mobility. Second- through fifth-grade data ($N = 1{,}003$, predominantly Caucasian) were drawn from a longitudinal study, and growth curve analyses allowed for the examination of mobility effects within the context of other factors that put children at risk, including behavior problems and family stress. School changes predicted declines in academic performance and classroom participation but not positive attitude toward school. Time-varying factors such as peer acceptance and teacher support had a positive influence on the growth trajectories of child outcomes. Additionally, teacher support had a particularly strong influence on positive attitudes toward school among children who had more school changes.

Educators and parents have long voiced concern about the effects of school enrollment changes on children's adjustment and achievement (Blane, Pilling, & Fogelman, 1985; Cramer & Dorsey, 1970), and accordingly, the consequences of student mobility have been the focus of numerous studies. A clear finding across this literature is that student mobility, defined as making a school enrollment change at a time not required by the grade structure of the school system, is associated with numerous other risk factors such as poverty, stressful life events (e.g., divorce), poor initial school performance, and a tendency to make additional enrollment changes during subsequent school years (Alexander, Entwisle, & Dauber, 1996; Eckenrode, Rowe, Laird, & Brathwaite, 1995; Kerbow, 1996; Nelson, Simoni, & Adelman, 1996; Pribesh & Downey, 1999). Given the intertwined nature of these factors, it can be difficult to isolate the impact of mobility from other factors that put children at risk for school problems. In a seminal article, Schaller (1976) cautioned researchers to control for preexisting differences when examining the effects of mobility, but not all subsequent researchers have heeded his warning.

Among the studies that have used controls, results suggest that student mobility has a negative effect on school performance, above and beyond the impact of other stressful features of a child's life (Astone & McLanahan, 1994; Haveman, Wolfe, & Spaulding, 1991; Heinlein & Shinn, 2000; Ingersoll, Scamman, & Eckerling, 1989). Some of the negative consequences associated with student mobility include: lower math and reading test scores (Mantzicopoulos & Knutson, 2000; Texas Department of Education, 1997), an increased risk of behavior problems (Tucker, Marx, & Long, 1998; Wood, Halfon, Scarlata, Newacheck, & Nessim, 1993), and a higher likelihood of being held back a grade (Simpson & Fowler, 1994; Tucker et al., 1998). Student mobility has also been shown to impact school completion and expected educational attainment (Astone & McLanahan, 1994; Hagan, MacMillan, & Wheaton, 1996; Pribesh & Downey, 1999; Rumberger & Larson, 1998; South, Haynie, & Bose, 2007).

Although a preponderance of the mobility research, and the related area of school transitions research (i.e., the study of scheduled enrollment changes to kindergarten, middle, or high school), has focused on either the experience of adolescents or the transition to kindergarten, the current study examines effects of mobility during the elementary school period. The gap in our knowledge is noteworthy because school mobility during the elementary

---

This study was supported by a grant from the National Institute on Drug Abuse (DA08093-13). The Raising Healthy Children project was conducted by the Social Development Research Group, which is part of the University of Washington School of Social Work. The authors gratefully acknowledge the staff, families, and students of Edmonds School District No. 15 for their support and cooperation. The lead author would also like to thank Susan Nolen and her colleagues in the Department of Educational Psychology at the University of Washington.

Correspondence concerning this article should be addressed to Diana H. Gruman, Department of Psychology, Western Washington University, 516 High Street, Bellingham, WA 98225-9089. Electronic mail may be sent to diana.gruman@wwu.edu.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7906-0021

school years may be strongly linked to long-term measures of school success, such as high school completion (Alexander et al., 1996; Haveman et al., 1991; Heinlein & Shinn, 2000).

How might enrollment changes made during elementary school influence long-term outcomes? Dropping out of school has been described as a developmental process of disengagement that may have its roots in the earliest grades (Finn, 1989; Rumberger, Larson, Ream, & Palardy, 1999). Research has shown that a scheduled transition to middle school or high school represents a dramatic environmental shift and a potential misfit for students that can begin to deteriorate some students' commitment to and sense of belonging at school (Alspaugh, 1998; Eccles & Midgley, 1990; Seidman & French, 1997). For elementary-age students making unscheduled school changes, this deterioration may be intensified because they typically are not involved in the decision to move and often face the adjustment process without the benefit of planned orientations (Jason et al., 1992). Additionally, researchers have concluded that elementary school mobility may have particularly harmful effects on academic performance because moving disrupts the acquisition of the critical building blocks on which later learning is based (Kerbow, 1996; Lash & Kirkpatrick, 1990).

In the present study, we sought to expand our understanding of how mobility during the elementary school years may contribute to the process of disengagement with school just prior to the significant changes that await children in adolescence. We wanted to improve on prior research and found the developmental science framework (Magnusson & Cairns, 1996) to be an instructive guide in determining our approach. The framework emphasizes that dynamic change occurs when individuals face environmental change, and it is therefore important to study adaptations to change "moment to moment," as the individual interacts with new features of the environment. Capturing the "emergence, dynamics, and pathways of change" in individual development (Magnusson & Cairns, 1996, p. 9) requires an analysis of longitudinal data. Typically, mobility researchers have used cross-sectional and retrospective designs and examined endpoint measures such as high school completion rates. In the context of this study, we chose to examine yearly measures of engagement and school experiences to allow us to consider temporal changes in exacerbating, confounding, and protective variables. We also determined that a multilevel analysis of the data would provide the best means to consider stable child characteristics as well as time-varying covariates that may impact individual students at each occasion of a school transfer (Raudenbush & Bryk, 2002; Snijders & Bosker, 1999).

Within the developmental science framework, it is recommended that researchers view the individual within a complex set of systems, taking into account the reciprocal process of an individual's interaction with the environment. Specifically, it is proposed that "variables in the developmental processes of individuals cannot be effectively studied if they are divorced from the social and environmental contexts in which they occur" (Magnusson & Cairns, 1996, p. 13). In the current study, we examine data from child, family, and school systems and focus attention on how school engagement outcomes are influenced by factors present in the school environment.

Among the challenges of conducting the type of research outlined in the developmental science framework is the retention of subjects (Magnusson & Cairns, 1996). Indeed, a number of the previous longitudinal studies of at-risk and mobile elementary-age students have lost between 20% and 25% of the original study population (see Alexander et al., 1996; Heinlein & Shinn, 2000; Temple & Reynolds, 1999). To ensure that the experiences of students who may be the most negatively affected by mobility were included, we used data from a study in which 94% of subjects were retained through the elementary grades (Fleming, Harachi, Catalano, Haggerty, & Abbott, 2001; Haggerty, Catalano, Harachi, & Abbott, 1998).

The Impact of Social Ties: Coleman's Theory of Social Capital and Vygotsky's Theory of Appropriation

One of the theories used commonly in the literature to explain the link between mobility and poor educational outcomes is the "social capital" theory of Coleman (1988, 1990). Closely related to the economic concepts of human and financial capital, social capital represents the relationship between the parents and the child and the network of relationships between parents, friends, and community members that may provide support to the family (Coleman, 1988). According to the theory, moving harms children's achievement because social ties are broken, thereby disrupting the exchange of social capital in the network. A number of controlled studies have drawn on Coleman's theory to explain how mobility, social capital, and achievement may be related (Hagan et al., 1996; Pribesh & Downey, 1999; South et al., 2007; Tucker et al., 1998). Because Coleman proposed a variety of indicators of social capital, each study uses a different measure of social capital.

In their study of mobility and high school dropout rates, Hagan et al. (1996) focused on the quality of the child's relationship with their parents and the father's level of participation in the family. They determined that "mother's support and father's participation are important sources of social capital that can mitigate the disruptive effects of family migration" (p. 381). In an attempt to broaden the definition to include social ties outside the family, Pribesh and Downey (1999) used six different measures of social capital, including participation in high school extracurricular activities, the frequency with which students discuss course planning with peers or parents, and the amount of contact parents have with other parents and school personnel. Moves were associated with declines in nearly all these measures of social capital, although higher levels of social capital appeared to buffer the negative effect of moving on test scores and educational attainment goals. Pribesh and Downey were also concerned about the impact of family stress experiences as a proxy for diminished social support. They determined that students whose parents divorced or who had one parent marry or die experienced declines in test scores independent of other variables. Finally, South et al. (2007) examined parent–child social ties, as well as other factors that might explain the higher dropout rates among mobile high school students, including psychological well-being (e.g., depression, self-esteem), school engagement, and peer friendships. They determined that for adolescents, lower levels of "peer network structure," measured by a student's place in a peer network and the density of the structure, provided the best explanation of movers' higher dropout rates.

This body of research has expanded our understanding of student mobility in a number of ways. First of all, most of these authors attended to both risk and protective factors in testing the impact of mobility. The typical approach has been to focus on how the severing of ties to familiar people and routines may negatively impact students. Research has shown, however, that not all highly mobile children suffer negative consequences (Kerbow, 1996; Mantzicopoulos & Knutson, 2000; Tucker et al., 1998). One result of the social capital research is that we are beginning to understand more about the factors that protect mobile children from harm following a school change. In particular, social support from family and peers appears to be powerful moderators of outcomes for mobile students. Still absent from the literature is a comprehensive study of the role that teachers play in supporting mobile children.

Additionally, by considering different spheres of influence surrounding the child and defining social capital in multiple ways, the authors who have examined social capital have tried to illuminate the complex relationship between mobility and educational outcomes. This approach has led to greater understanding, for example, of the role of stressful family events, such as divorce, that may operate independently of changing schools to erode the support that children need to succeed in school. Unfortunately, the lack of definitional clarity has been one of the criticisms of social capital theory in recent years (see Durlauf, 1999; Morrow, 1999), leading one author to conclude that social capital is better construed as a "descriptive construct rather than an explanatory model" (Morrow, 1999, p. 760). In the context of the present study, we determined that a test of "social ties" within the school environment was more appropriate than a test of social capital per se. Here, we are interested in how social ties with teachers and peers in the school environment serve to support successful outcomes for students experiencing family stress and/or mobility.

Complementing the perspective of social capital theory, Vygotsky's (1962, 1978) sociocultural theory of "appropriation" describes the processes through which supportive teacher and peer relationships may provide an important positive influence for mobile children. Summarized by Thompson (1998), the model of appropriation "underscores how social interaction provides the medium by which socioemotional skills, understanding, and perspective are jointly created as the child participates with others in shared activity, whether consisting of face-to-face exchanges, collaborative play, or the soothing of fears or distress" (Thompson, 1998, p. 27).

Important from a mobility perspective, Vygotsky emphasized that adaptation to the environment is achieved by social means (Goldstein, 1999; van der Veer & Valsiner, 1994; Wertsch, 1985). When we consider the experience of the child who has just changed schools, perhaps overwhelmed by the new physical space and mourning the loss of friends in a previous school, we expect that there may be a period of adjustment during which time the child may not show signs of academic progress (Kerbow, 1996; Vernberg, 1990). Once past the orientation stage (e.g., becoming acquainted with the schedule of daily activities and academic/behavioral expectations), we hope that a child will not only adapt to the new setting but also begin to develop social and academic skills at the same rate as before moving. Drawing on the theory of appropriation, what creates a bridge between bewilderment and success in the new setting is helpful social contact with a teacher and peers.

Empirical evidence to support the role of social contacts in mitigating stressful life events for children is abundant (DuBois, Felner, Meares, & Krier, 1994;

Garmezy, 1983; Hendershott, 1989; Sandler, Miller, Short, & Wolchik, 1989; Wenz-Gross, Siperstein, Untch, & Widaman, 1997), with research indicating that strong connections among students, teachers, and positive peers facilitate access to social and scholastic activities (Hughes, Cavell, & Willson, 2001; Ladd, Birch, & Buhs, 1999; Ladd & Burgess, 2001; Wentzel, 1997). Research has also shown that student–teacher relationships characterized as supportive, close, or caring, have significant effects on academic motivation, attitude, and engagement in the school environment (Birch & Ladd, 1997; Wentzel, 1997).

In the area of student mobility, studies of the scheduled transition to middle school or high school indicate that support from peers and teachers positively influences the academic and social adjustment of adolescents (Felner, Ginter, & Primavera, 2002; Hirsch & DuBois, 1992; Isakson & Jarvis, 1999; Lord, Eccles, & McCarthy, 1994). As we have mentioned, less is known about the relational factors that influence the school success of elementary school students making unscheduled school changes. For this study, we are interested in the potential protective qualities of peer and teacher relationships for these mobile students.

School Engagement Outcomes and Mobility

What measures of school engagement do we think might be negatively influenced by mobility and positively influenced by supportive peer and teacher relationships? Recognizing that a number of key shifts in children's attitudes and behaviors may occur when they experience enrollment changes in the early years of schooling, we selected three dimensions of engagement to shed light on the academic, affective, and behavioral facets of school success (Finn, 1989; Finn & Rock, 1997). First of all, we include a measure of teacher-rated academic performance in three core curricular areas—reading, math, and language arts—to confirm the results of previous studies (Kerbow, 1996; Tucker et al., 1998). Second, we include a measure of children's attitudes toward school to tap the affective dimension of school engagement (Alexander, Entwisle, & Horsey, 1997). We included an affective measure because mobility has been linked with increased rates of behavior problems (e.g., suspension, disobedience) in elementary-age students (Tucker et al., 1998; Wood et al., 1993), and though as yet untested, we suspect that the increased difficulties that mobile students experience might erode the positive attitudes students hold about school.

Finally, the development of positive classroom behaviors such as participation and cooperation skills may be compromised when children, lacking consistent reinforcement and support, move from setting to setting. Though untested in the context of mobility, we believe that these "deportment" behaviors are worth examining given the empirical support for their influence on academic performance (Alexander, Entwisle, & Dauber, 1993; Alexander et al., 1997; Ladd et al., 1999). We therefore include a scale called "classroom participation" that indicates the extent to which an individual interacts appropriately with teachers and peers in the classroom.

Current Study

The purpose of this study is to open up a more dynamic view of mobility than has previously been attempted, using longitudinal data to explore how mobility during the elementary school years might undermine or erode the skills and attitudes that typically lead to successful school outcomes. Data from second through fifth grades were drawn from an ongoing longitudinal study in which researchers successfully retained a high percentage of subjects (94%) despite a high degree of mobility within the subject pool (Fleming et al., 2001; Haggerty et al., 1998). Within the design, we considered how risk factors as well as protective factors in the school and home environment might influence academic and behavioral outcomes at fifth grade (see model design in Figure 1).

Children who move, especially frequently, also tend to experience other stressors that make them more vulnerable to school failure. Thus, we included measures of additional risk factors (represented on the left side of the model in Figure 1) in our tests of

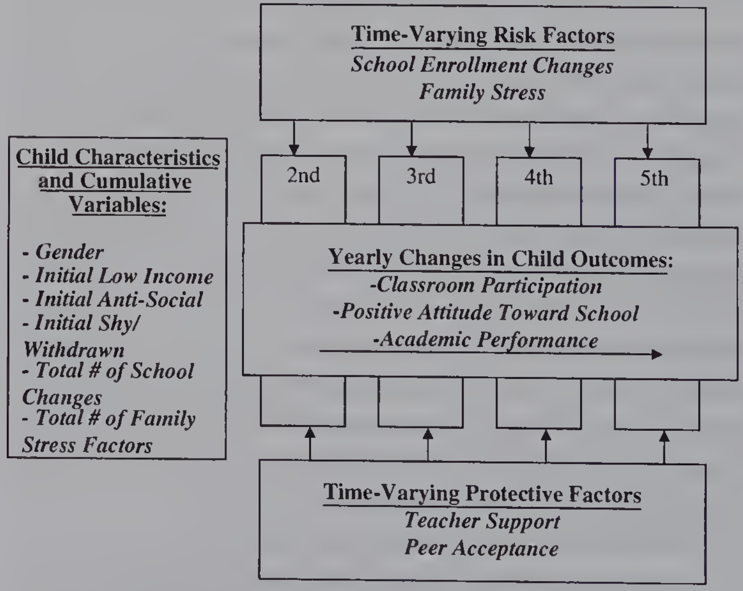

Time-Varying Risk Factors
School Enrollment Changes
Family Stress
Child Characteristics and Cumulative Variables:
- Gender
- Initial Low Income
- Initial Anti-Social
- Initial Shy/ Withdrawn
- Total # of School Changes
- Total # of Family Stress Factors
2nd
3rd
4th
5th
Yearly Changes in Child Outcomes:
-Classroom Participation
-Positive Attitude Toward School
-Academic Performance
Time-Varying Protective Factors
Teacher Support
Peer Acceptance

Figure 1. Model of the relationship between outcomes, time-varying covariates, and child characteristics.

mobility to understand the stressful context of mobile children's lives. Low socioeconomic status and gender (being male) are included because they have been shown to increase the risks of negative school outcomes for mobile children (Ellickson & McGuigan, 2000; Heinlein & Shinn, 2000; Jason et al., 1992; Mantzicopoulos & Knutson, 2000; Nelson et al., 1996; Tucker et al., 1998; Wood et al., 1993). Initial shy/withdrawn behavior and initial antisocial behavior are included because they have been tied to poor educational outcomes and may hinder the development of relational supports available to mobile students (Coie, Dodge, & Kupersmidt, 1990; Ladd & Burgess, 2001; Wentzel, 1993). We further recognize that when students face chronic stress or concurrent stresses, they have a more difficult time coping (Hetherington, 1979; Hofferth, Boisjoly, & Duncan, 1998; Moore, Vandivere, & Ehrle, 2000; Simmons, Burgeson, Carlton-Ford, & Blyth, 1987; Tucker et al., 1998). We therefore include two variables, total school changes and total stressful life events, to test the cumulative effects of these experiences.

In addition, we included time-varying factors measured at each grade level to determine whether dynamic factors play a role in changing the trajectories of growth for mobile students (Magnusson & Cairns, 1996). Data for these four factors, gathered through child, parent, and teacher surveys, include two risk factors: children's yearly experiences with family stress and school enrollment changes (represented in the top box of the model in Figure 1) and two protective factors: peer acceptance and teacher support (represented in the lower box in Figure 1). The three outcome measures (in the center boxes in Figure 1), representing different behavioral and attitudinal components of school engagement, were taken from teacher and child surveys. These include a measure of teacher-rated academic performance, a measure of classroom participation that indicates that an individual is participating successfully in the academic and social life of the classroom, and a measure of positive attitude toward school to determine how mobility impacts the affective ties that children have toward school. The method of analysis, linear growth curve modeling, allowed us to examine the interactions between the individual, their family circumstances, and their school relationships across time, thereby modeling the complexity of the mechanisms through which mobility may impact school performance and related skills and attitudes (Adelman & Taylor, 1991; Ladd & Burgess, 2001; Reynolds & Bezruczko, 1993).

In summary, this investigation seeks to assess both the short-term effects of mobility, through yearly measures, and trends over multiple years, though cumulative measures. We are further interested in the role that family stress (measured both yearly and cumulatively), teacher support (measured yearly), and peer acceptance (measured yearly) play in influencing student engagement. We sought to answer three questions: First, do experiences with elementary school mobility, measured cumulatively, diminish attitudes toward school and reduce the development of classroom participation and academic performance? Second, do risk factors, including gender, low-income status, antisocial and shy/withdrawn behavior, and overall family stress, play a role in explaining the relationship between mobility and school outcomes? Third, do yearly measures of school change, family stress, peer acceptance, and teacher support impact the development of social and academic outcomes, taking into account other risk factors?

Method

Raising Healthy Children Project

This study analyzed data from the Raising Healthy Children (RHC) project, a prospective longitudinal intervention study conducted by the Social Development Research Group. In the fall of 1993, 10 public elementary schools in a Pacific Northwest suburban school district were selected because they had student populations with high aggregate measures of risk relative to other schools in the district. Risk was based on the percentage of students having below-average standardized reading scores, living in single-parent households, receiving subsidized school lunches, and experiencing high rates of mobility. Eligibility criteria for the sample included: attending a regular education first- or second-grade classroom, remaining in the project school from the time of parental consent (January 1994) through to completion of the first student survey in the spring of 1994, and having a parent who spoke English, Spanish, Vietnamese, or Korean.

Parents of 938 (76%) of the 1,239 eligible students consented to participation. In the 2nd year of the study, second- and third-grade students who had newly enrolled into a project school were also invited to participate in the study, bringing the overall study cohort to 1,040 students. There were no significant differences on basic demographic characteristics between consenting and nonconsenting families. Data in the RHC project were collected on a yearly basis from district records and from annual parent, child, and teacher surveys administered each spring. Parents received a $10 incentive for survey participation, students received a small gift, and teachers received a $5 incentive for the assessment of each consenting student in their classroom. Particular attention was paid to retaining a maximum number of participants, even if they moved out of the state. A 94% retention rate was maintained at the end of the 5th year of data collection, when students were enrolled in fifth or sixth grade.

RHC is a study of the etiology of problem behaviors as well as a randomized test of a multicomponent preventive intervention. The intervention was delivered at five of the project elementary schools and consisted of instructional staff development for teachers, parenting workshops, booster sessions for parents and teens, summer camps and study clubs for students, and home-based case management services for high-risk students who exhibited academic or behavioral problems. Additional details regarding the RHC intervention have been reported by Catalano et al. (2003), Haggerty et al. (1998), and Brown, Catalano, Fleming, Haggerty, and Abbott (2005).

RHC is an ongoing longitudinal intervention study. An experimental/control variable was included originally in the analytic plan for the present study. In this analysis, experimental effects on the outcomes were examined using both correlational methods and hierarchical linear modeling, but at no point during the analytic process was an experimental effect detected. Therefore, data in this present study were pooled and include children in both the experimental and the control conditions.

Sample

Data on students when they were enrolled in Grades 2 through 5 were used for this study. Of the 1,040 children originally enrolled in the RHC project, 37 were removed from the current analysis. First, 9 students who were homeschooled during the elementary grades were eliminated because they had limited teacher data and lacked school-based experiences important in this study. Second, the statistical design required a complete set of time-invariant individual data for each subject, which reduced the study by another 28 students, leaving a total of 1,003 for the present analysis.

The sample contained roughly equal numbers of boys and girls (471 girls and 532 boys), and 35% of parents reported that their families were low income during one of the first 2 years of the study ($n = 355$). In terms of mobility, nearly 50% of children remained in the same school during second through fifth grades ($n = 497$), 33% made one school change ($n = 328$), 13% made two changes, and 5% of the sample made three or more school changes ($n = 47$). For stressful family events, characterized by changes in parental relationships, family experience with job loss, serious accidents or illnesses, or death of a household member, 48% of children ($n = 478$) experienced either zero or one stressful event over the 4-year study period. Twenty-eight percent experienced two or three events ($n = 276$), 14% experienced four or five events ($n = 143$), and 9% ($n = 90$) experienced six or more stressful family events.

In terms of ethnic group composition, the sample represented a population typical of many suburban school districts: 82% ($n = 818$) European American, 7% ($n = 72$) Asian/Pacific Islander, 4% ($n = 43$) Hispanic, 4% ($n = 43$) African American, and 3% ($n = 27$) Native American. The lack of diversity in the sample limited our analysis of race/ethnicity in this study. However, we did examine whether mobility, family stress, and other risk/protective factors were associated with the available categories of race/ethnicity and found that there were no significant unique associations after controlling for low-income status, gender, antisocial, and shy/withdrawn behavior.

Measures

In the following list of measures, items taken from existing measures cite the source; otherwise, the items were created for use in the RHC project. Means, standard deviations, alpha reliabilities, data source, and data collection points are provided in Table 1.

Outcomes

Classroom participation. A nine-item scale assessing classroom participation was drawn from teacher surveys. The scale consists of items derived from the Walker – McConnell Scale of Social Competency and School Adjustment (Walker & McConnell, 1988) and Teacher Observation of Classroom Adaptation – Revised (TOCA – R; Werthamer-Larsson, Kellam, & Ovesen-McGregor, 1990). Yearly alpha reliabilities for Grades 2 through 5 were .91, .90, .90, and .90. Example items include: (a) interacts appropriately with teacher, (b) cooperates with peers in group activities, and (c) listens while others are speaking. The response options for items ranged from either *rarely or never true* to *often true* or *not at all* to *often true*. A mean score was computed across items so that the scale had a range from 1 to 5.

Table 1

Descriptive Statistics for Hierarchical Linear Modeling Analyses

Variable	*M*	*SD*	*N*	α	Data source	Data collection
Time invariant						
Gender	0.53	0.50	1,003		Parent	Year 1 or 2
Low income	0.35	0.48	1,003		Parent	Year 1 or 2
Antisocial	1.24	0.42	1,003	.88	Teacher	Year 2, fall
Shy/withdrawn	1.34	0.40	1,003	.81	Teacher	Year 2, fall
Total school changes	0.74	0.90	1,003		School	Yearly, spring
Total stress	2.14	2.18	1,003		Parent	Yearly, spring
Time varying						
Yearly school change	0.18	0.39	3,917		School	Yearly, spring
Yearly stress	0.58	0.79	3,684		Parent	Yearly, spring
Peer acceptance	0.00	0.85	3,759	.75–.89	Teacher	Yearly, spring
Teacher support	3.46	0.54	3,827	.54–.78	Child	Yearly, spring
Classroom participation	3.80	0.78	3,755	.90–.91	Teacher	Yearly, spring
Attitude toward school	3.15	0.62	3,831	.55–.70	Child	Yearly, spring
Academic performance	3.19	1.23	3,753	.92–.93	Teacher	Yearly, spring

Note. The time-invariant variables were entered at Level 2. The time-varying variables were entered at Level 1.

Positive attitude toward school. A four-item scale assessing a child's attitude toward school was drawn from child surveys administered each year. Items were taken from the Seattle Social Development Project (Hawkins & Catalano, 1990) and the Seattle Personality Questionnaire (SPQ; Kusché & Greenberg, 1988). Items include: (a) Do you think school is fun? (b) Do nice things happen to you at school? (c) Do you feel unhappy at school a lot? (d) Do you look forward to going to school? Response options were *YES, yes, no, NO.* A mean score was computed across items so that the scale had a range from 1 to 4. Though internal consistency was moderate, alpha reliabilities increased as children grew older (α = .55, .62, .70, and .70).

Academic performance. Teacher ratings of a child's academic performance in three content areas (language arts, math, and reading) were collected each year. The 5-point response options for each item were *above average, slightly above average, average, slightly below average,* and *needs improvement.* A scale was created by averaging the three academic areas each year with a range of 1–5. Due to the frequent school changes within the study population (across regions and states), we are unable to use standardized test data without compromising the sample retention rate. However, construct validity of our measure of academic performance is supported by a .79 correlation between this measure and a composite of math and reading scores on the California Test of Basic Skills administered to the students who were still enrolled in the same school district in fourth grade.

Time-Varying Measures

Yearly school changes. A school transfer variable was coded in each of the 4 years as 0 = *no school transfer* and 1 = *one or more school transfers.* Transfers were calculated for the period between June of one school year and June of the next school year using district and school office records.

Yearly family stress. Stressful events in the life of each child were assessed yearly through parent surveys and were summarized into a 0–4 index. The events included: changes in parental relationships, family experience with job loss, serious accidents or illnesses, and death of a household member over the course of the past year.

Teacher support. Four items assessed a child's perception of teacher support. Items were drawn from the Seattle Social Development Project (Hawkins & Catalano, 1990) and were designed to measure prosocial processes in the classroom (Catalano et al., 2005). The items include: (a) Does your teacher notice when you try hard? (b) Does your teacher tell you when you do well? (c) Does your teacher help you when you need it? (d) Do you like your teacher? Response options were *YES, yes, no, NO.* A mean score was computed across items so that the scale had a range from 1 to 4. Though internal consistency was moderate, alpha reliabilities increased as children grew older (αs = .54, .68, .75, and .78).

Peer acceptance. A three-item scale representing level of peer acceptance was drawn from teacher surveys administered each spring. Items were taken from the Walker T Pref. (Walker & McConnell, 1988) and from the TOCA–R (Werthamer-Larsson et al., 1990). The items include: (a) makes friends easily with other children, (b) plays with classmates, and (c) rejected by classmates (reversed). The number of response options differed across the items; therefore, a scale score was created by first computing a z score for each item and then averaging across the three z scores.

Measures of Individual Characteristics and Cumulative Measures of Stress and Mobility

Gender. Gender data were taken from district records and coded as a dichotomous variable, with 0 = *female* and 1 = *male.*

Initial low income. Initial low-income status was assessed during the first 2 years of the project through parent surveys and was coded as a dichotomous variable with 0 = *not low income* and 1 = *low income.* Families who reported being eligible for free or reduced-price lunch program, Aid to Families with Dependent Children, or food stamps during Year 1 or Year 2 of data collection were coded as having initial low-income status.

Initial antisocial behavior. A five-item scale of antisocial behavior was drawn from teacher surveys administered in the fall when the children were in second or third grade. The scale consists of items taken from the TOCA–R (Werthamer-Larsson et al., 1990) and the Child Behavior Checklist (Achenbach, 1991). Items include: (a) argues a lot, (b) talks back to adults/is disrespectful, (c) lies, (d) stubborn, and (e) yells at others. Response options were: *rarely or never true*, *sometimes true*, and *often true*. A mean score was computed across items so that the scale had a range from 1 to 3. Correlations across Grades 2 through 5 ranged from .43 to .62 ($p < .001$).

Initial shy/withdrawn behavior. A five-item scale of shy/withdrawn behavior was drawn from teacher surveys administered in the fall when the children were in second or third grade. The scale consists of items taken from the TOCA–R (Werthamer-Larsson et al., 1990) and the Child Behavior Checklist (Achenbach, 1991). Items include: (a) shy or timid; (b) withdrawn, does not get involved with others; (c) avoids classmates; (d) likes to be alone; and (e) secretive, keeps things to self. Response options were: *rarely or never true*, *sometimes true*, and *often true*. A mean score was computed across items so that the scale had a range from 1 to 3. Correlations across Grades 2 through 5 ranged from .27 to .43 ($p < .001$).

Total school changes. A cumulative index of school changes was calculated by summing the total number of school changes a child experienced during the 4 study years. This measure included multiple school changes that some subjects made within one school year: Thirty-one students moved twice in 1 year and 2 students moved three times in 1 year. The most mobile students in the study experienced five school changes; thus, the variable ranged from 0 to 5.

Total family stress. A cumulative index of family stress was calculated by summing children's yearly experiences with family stress. Because four items were assessed during each of the 4 study years, students could have experienced a maximum of 16 stressors. The actual range for this sample was 0–13.

Analysis

Growth curve analyses were conducted using hierarchical linear modeling (HLM) 5.0 (Raudenbush, Bryk, Cheong, & Congdon, 2000). In the growth curve analysis framework, individual change is represented through a two-level hierarchical model in which the repeated measurements at Level 1 (e.g., yearly values on the outcome measures or covariates) are nested within individuals at Level 2 (e.g., person-level predictors such as gender and low-income status). In addition to the intercept and linear change parameters in the general Level 1 model, the time-varying covariates (yearly experiences with school transfers, teacher support, peer acceptance, and family stress) were specified at Level 1. Child characteristics (gender, initial low-income status, and initial antisocial and shy/withdrawn behavior) and cumulative variables (total number of school transfers and total number of family stressors) were included in the Level 2 equations.

Selecting the "center point" for time is a critical decision in the HLM process because it determines how one will interpret individual coefficients. Time in this study was coded (−3, −2, −1, 0) so that the interpretation of the intercept was at the last time point when the children were in fifth grade. The intercept was defined as the outcome at fifth grade because we were interested in examining specifically if school changes and other life circumstances altered the level of outcomes by the last time point in the study. Additionally, setting the intercept at the last time point ensured temporal ordering between the intercept and the cumulative school change and cumulative family stress variables.

Results

Analysis of the data proceeded in four phases. The first phase involved examining correlational data to consider the relationships among the variables prior to modeling growth in the outcome variables. In the next three phases of analyses, HLM models were built using a "step-up" strategy, starting by testing simple relationships and moving to more complex models with multiple variables (Raudenbush & Bryk, 2002). We selected a step-up strategy because in our initial analyses we tested models with all Level 1 and Level 2 predictors simultaneously and found that we did not have enough power to test these complex models.

Phase 1: Correlations

The matrix of correlations for child characteristics and across-time averages of each of the time-varying covariates and outcomes is provided in Table 2. Among the findings, the total number of school changes was positively associated with being low income at the beginning of the study and with receiving a higher teacher rating of antisocial behavior in the early years of data collection. School changes were also positively associated with the total number of stressful life events a child experienced during the study period. Finally, in terms of its relationship with the outcome measures, total school changes were correlated with academic performance and classroom participation but not with positive attitude toward school.

Another noteworthy finding was that, contrary to our hypothesis, teacher support was not correlated significantly with academic performance. Finally, across-time peer acceptance was found to be highly correlated with across-time classroom participation ($r = .67, p < .001$). Both of the measures were based on teacher ratings, and there was a concern that they were measuring the same construct. Therefore, peer acceptance was excluded in the models examining classroom participation and results do not appear in subsequent tables.

Phase 2: Fitting the Level 1 Models

A series of models for each of the three outcome variables examined the intercept (grand mean at fifth grade), the linear change (rate of growth), and slopes for each of the four time-varying predictors. Results of this phase of the analysis are shown in Table 3. First, an unconditional model containing the intercept and the linear change slope was estimated for each outcome (Table 3, column 1). The model provides a baseline measure of the intercept and average rate of growth on each outcome. Results indicate that each of the three outcomes was characterized by significant negative linear change.

Table 2
Correlations

	1	2	3	4	5	6	7	8	9	10
1. Total school changes										
2. Gender	.02									
3. Low income	.22**	.01								
4. Antisocial	.13**	.15**	.14**							
5. Shy/withdrawn	.07**	–.02	.11**	.16**						
6. Total stress	.17**	.02	.26**	.14**	.05**					
7. Yearly school change	.54**	.01	.12**	.07**	.04**	.08**				
8. Yearly stress	.13**	.01	.20**	.11**	.03*	.71**	.07**			
9. Peer acceptance	–.12**	–.10**	–.17**	–.34**	–.26**	–.14**	–.08**	–.11**		
10. Teacher support	.04**	–.01	.04**	–.09**	.00	–.01	.03	.02	.08**	
11. Classroom participation	–.13**	–.22**	–.18**	–.42**	–.14**	–.15**	–.08**	–.12**	.67**	.15**
12. Attitude toward school	.01	–.09**	–.02	–.16**	–.01	–.03	.00	–.01	.22**	.55**
13. Academic performance	–.10**	–.06**	–.25**	–.17**	–.13**	–.11**	–.08**	–.09**	.36**	.01

Note. Correlational data for the time-varying measures, including yearly school change, yearly stress, peer acceptance, teacher support, classroom participation, attitude toward school, and academic performance represent across-time averages.
*Pearson *r* correlation is significant at the .05 level. **Pearson *r* correlation is significant at the .01 level.

Table 3

Comparing Alternative Level 1 Models

	Unconditional		Unconditional with yearly school change		Unconditional with yearly stress		Unconditional with peer acceptance		Unconditional with teacher support	
	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*
Classroom participation										
Intercept at Grade 5	3.736***	.025	3.737***	.025	3.735***	.025			3.736***	.025
Linear change	−0.038***	.009	−0.038***	.009	−0.039***	.009			−0.038***	.009
Covariates										
Yearly school change			−0.026	.028						
Yearly stress					−0.011	.017				
Peer acceptance										
Teacher support									0.117***	.021
Positive attitude										
Intercept at Grade 5	3.099***	.200	3.098***	.200	3.101***	.019	3.099***	.200	3.098***	.018
Linear change	−0.032***	.009	−0.033***	.009	−0.031***	.009	−0.032***	.009	−0.032***	.007
Covariates										
Yearly school change			.015	.027						
Yearly stress					0.019	.017				
Peer acceptance							0.064***	.018		
Teacher support									0.578***	.019
Academic performance										
Intercept at Grade 5	3.113***	.039	3.120***	.039	3.112***	.039	3.117***	.039	3.113***	.039
Linear change	−0.046***	.011	−0.042***	.011	−0.046***	.011	−0.043***	.011	−0.046***	.011
Covariates										
Yearly school change			−0.146***	.037						
Yearly stress					−0.008	.021				
Peer acceptance							0.222***	.022		
Teacher support									0.048	.028

***$p < .001$.

Next, the four time-varying predictors (yearly measures of school change, experience with family stress, teacher support, and peer acceptance) were added to the unconditional growth models one at a time to estimate the effects of children's exposure to these stressful and supportive conditions (Table 3, columns 2–5). The only exception, as mentioned earlier, was that the model testing the influence of peer acceptance on classroom participation was excluded due to high correlations. Each time-varying predictor was centered in a theoretically meaningful way to aid in the interpretation of the results. First, for yearly transfers and yearly stress, 0 represented the absence of a transfer or no experience with any of the four stressful family events. Second, the teacher support and peer acceptance predictors were "group-mean centered." Group, in this context, refers to the group of scores available for each child, and thus, within the analysis, the impact of teacher support and peer acceptance is in relationship to a student's own experiences over time not to that of his or her peers. Our choice to use group-mean centering stemmed from a concern that students who typically receive more support are the ones who do better in school. The correlation between across-time teacher support and across-time classroom participation ($r = .15$, $p < .001$) and positive attitude toward school ($r = .55$, $p < .001$) supports this relationship.

Results for the time-varying predictors indicate that peer acceptance and teacher support were significant and positively related to the prediction for two of the three tested outcomes (Table 3, columns 2–5). In addition, yearly transfers were significantly and negatively related to teacher ratings of academic performance. A child's yearly experience with stress was not a significant predictor for the growth trajectories of any of the three outcomes. Only the significant time-varying predictors were selected to move into the next phases of the model-building process.

A final step in this phase of the analysis involved testing cross-level interactions between the Level 2 child predictors and the slopes for time and each of the significant time-varying covariates. This step was undertaken to determine whether certain characteristics of the child were related to accelerations or decelerations in linear growth. To test these relationships, all Level 2 child predictors were tested on the slope of time and on the significant covariates for each outcome. Three significant or nearly significant cross-level interactions emerged from these tests and were included in subsequent tests to determine if they retained significance in the more complex models. These results are not shown in the tables. First, in the model testing the effect of teacher support on positive attitude toward school, the total number of school changes was significantly related to teacher support (coefficient = .055, $SE = .022$), $t(986) = 1.97$, $p < .021$. Second, for academic performance, two predictors approached significance: The total number of school changes slowed the growth of academic performance over time (coefficient = −.027, $SE = .013$), $t(986) = -2.34$, $p < .051$, and initial antisocial behavior predicted a reduction in the effect of peer acceptance on academic performance (coefficient = −.092, $SE = .046$), $t = -2.08$, $p < .056$.

Phase 3: Adding Level 2 Predictors and Significant Level 1 Covariates

This phase of analysis answered the first two research questions proposed for this study. The results are summarized for each outcome separately in Tables 4–6. The first question was, "Do experiences with elementary school mobility, measured cumulatively, diminish attitudes toward school and reduce the development of classroom participation and academic performance?" To answer this question, the total number of school changes a child made over 4 years was added to the intercept and linear change parameters within the unconditional growth model (Tables 4–6, column 1). Total school changes had a significant negative impact on fifth-grade classroom participation scores, $t(986) = -4.24$, $p < .001$, and scores for academic performance, $t(986) = -4.18$, $p < .001$, but was not significant for a child's positive attitude toward school. This is noteworthy because the experience of transferring schools was conceived as a risk factor, which might erode children's positive attitudes about school. These findings do not support this hypothesis. In terms of linear change, the total number of school transfers a child made significantly impacted the linear change over time for only one outcome: academic performance, $t(987) = -2.10$, $p < .05$. Because the number of transfers was significant in this step, it was retained as a predictor of the slope in subsequent models for academic performance. It was not retained as a predictor on the time slopes of the other two adjustment outcomes.

The second research question proposed in this study was, "Do risk factors, including gender, low-income status, antisocial and shy/withdrawn behavior, and overall family stress, play a role in explaining the relationship between mobility and school outcomes?" To answer this question, the remaining Level 2 child characteristics were included in the model on the intercept (Table 4–6, column 2). Results indicate that when the other five predictors are included in the

Table 4
Classroom Participation—Adding Level 2 Predictors and Significant Level 1 Covariates

	Unconditional plus total school changes		Unconditional plus all Level 2 predictors		Final model with teacher support	
	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*
Intercept at Grade 5	3.819***	.032	4.018***	.035	4.017***	.035
Total school changes	−0.115***	.027	−0.042*	.021	−0.042*	.020
Gender			−0.274***	.033	−0.277***	.033
Low income			−0.157***	.037	−0.153***	.037
Antisocial			−0.679***	.044	−0.683***	.045
Shy/withdrawn			−0.133**	.049	−0.129**	.049
Total stress			−0.022**	.008	−0.022**	.009
Linear change	−0.039***	.012	−0.038***	.009	−0.037***	.009
Total school changes	0.001	.011				
Covariate: Teacher support					0.118***	.020

*$p < .05$. **$p < .01$. ***$p < .001$.

models, the student mobility variable continues to have a significant negative effect on the intercept of classroom participation at fifth grade, $t(981) = -2.02$, $p < .05$, but not on academic performance. However, total transfers continue to significantly slow the linear growth of academic performance, $t(986) = -2.11$, $p < .05$.

Within these models, other child-level predictors did emerge as significant in predicting scores for the intercept at fifth grade. Initial antisocial behavior and low-income status contributed significantly to the intercepts of all three outcomes. Being male predicted declines in positive attitude toward school and classroom participation and having shy/withdrawn behavior predicted declines in classroom participation and academic performance. Finally, the total number of stressful family events a child experienced predicted declines in classroom participation.

Phase 4: Testing the Final Models

In the final stage of the analysis, each significant Level 1 covariate was entered one at a time to answer the third research question: "Do yearly measures of school change, family stress, peer acceptance, and teacher support impact the development of social and

Table 5
Positive Attitude Toward School—Adding Level 2 Predictors and Significant Level 1 Covariates

	Unconditional plus total school changes		Unconditional plus all Level 2 predictors		Final model with teacher support		Final model with peer acceptance	
	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*
Intercept at Grade 5	3.086***	.025	3.134***	.030	3.132***	.029	3.135***	.030
Total school changes	0.020	.022	0.005	.015	0.003	.015	0.005	.015
Gender			−0.081**	.027	−0.082**	.027	−0.082**	.027
Low income			0.058*	.029	0.055	.029	0.058*	.029
Antisocial			−0.236***	.038	−0.227***	.038	−0.235***	.038
Shy/withdrawn			0.016	.035	0.015	.035	0.016	.035
Total stress			−0.007	.006	−0.006	.006	−0.007	.006
Linear change	−0.043***	.012	−0.032***	.009	−0.033***	.007	−0.032***	.009
Total school changes	0.016	.011						
Covariate: Teacher support					0.543***	.025		
Covariate: Total school changes					0.045*	.023		
Covariate: Peer acceptance							0.063***	.018

*$p < .05$. **$p < .01$. ***$p < .001$.

Table 6

Academic Performance—Adding Level 2 Predictors and Significant Level 1 Covariates

	Unconditional plus total school changes		Unconditional plus all Level 2 predictors		Final model with peer acceptance		Final model with yearly school change	
	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*
Intercept at Grade 5	3.244***	.051	3.464***	.068	3.470***	.068	3.463***	.068
Total school changes	−0.184***	.044	−0.081	.021	−0.086	.044	−0.046	.045
Gender			−0.111	.033	−0.110	.066	−0.109	.066
Low income			−0.549***	.037	−0.551***	.073	−0.552***	.073
Antisocial			−0.308***	.044	−0.306***	.083	−0.309***	.084
Shy/withdrawn			−0.290***	.049	−0.290***	.083	−0.287***	.083
Total stress			−0.018	.008	−0.019	.016	−0.019	.016
Linear change	−0.027*	.014	−0.027	.014	−0.022	.013	−0.026	.014
Total school changes	−0.028*	.013	−0.028*	.013	−0.031*	.013	−0.022	.013
Covariate: Peer acceptance					0.230***	.020		
Covariate: Antisocial					−0.099*	.047		
Covariate: Yearly school change							−0.115***	.038

$^{*}p < .05$. $^{***}p < .001$.

academic outcomes, taking into account other risk factors?" Results of final models for each of the three outcomes are presented in Tables 4–6.

Classroom Participation

As shown in Table 4, column 3, teacher support had a positive and significant influence in the model, $t(981) = 5.65$, $p < .001$. Specifically, in years when children reported higher than average levels of teacher support, compared to their own scores across time, their classroom participation scores were estimated to improve. The total number of school changes retained significance, $t(981) = -2.03$, $p < .05$, even with the other five Level 2 predictors included in the model. Gender, initial antisocial behavior, and initial low-income status emerged as unique predictors of declines in children's classroom participation scores.

Positive Attitude Toward School

In the model testing the time-varying influence of teacher support (Table 5, column 3), teacher support was predictive of the growth trajectories for children's positive attitudes toward school, $t(986) = 22.13$, $p < .001$. In this model, gender and initial antisocial behavior were still significant predictors of positive attitude toward school, but low income was no longer a significant predictor.

The cross-level interaction between teacher support and the total number of school changes was also included in the analysis of attitude toward school (Table 5, column 3). Results indicate that the number of school changes a child experienced increased the strength of the relationship between teacher support and attitude toward school. In other words, teacher support is, in general, related to children's positive attitude toward school but is particularly important for those children who experience school changes.

The model including peer acceptance (Table 5, column 4) predicts a positive change in the growth trajectories for positive attitude toward school, $t(987) = 3.45$, $p < .001$. Of the child-level variables entered in the model, gender, initial antisocial behavior, and low-income status were significant predictors of fifth-grade positive attitude toward school.

Academic Performance

In the analysis of academic performance, results indicate that peer acceptance (Table 6, column 3) made a significant positive contribution, $t(986) = 10.23$, $p < .001$. In this model, three Level 2 predictors were statistically significant at $p < .001$ for academic outcomes at fifth grade: initial low-income status, antisocial behavior, and shy/withdrawn behavior.

Within the peer acceptance model, two cross-level interactions were significant. The interaction between peer acceptance and initial antisocial behavior was significant, $t(986) = -2.08$, $p < .05$, indicating that antisocial behavior decreases the effect of peer acceptance on academic performance. Additionally, the variable for total number of transfers, included on the time slope, was also significant, $t(986) = -2.34, p < .05$, indicating that the linear growth of academic performance was slowed by school transfers.

A second model, estimating the influence of yearly school changes on academic performance, was tested in a similar fashion (Table 6, column 4). In this model, yearly school changes predicted declines in the growth of academic performance, $t(987) = -3.02$, $p < .01$. Of the Level 2 child characteristics, three were significant at $p < .001$: initial low-income status, antisocial behavior, and shy/withdrawn behavior. Total number of transfers no longer predicted a change in the linear growth of academic performance.

Discussion

Research on mobility in schools has, until now, been limited by an inability to capture the complex conditions in which children make school changes. In this study, the application of growth curve analyses within a developmental science research framework allowed for the examination of school transfer effects within the context of other characteristics that may put children at risk for poor adjustment. Specifically, by including risk factors such as antisocial behavior, low-income status, and family stress, growth curve modeling enabled us to examine the unique effect of mobility. The analytic tools also allowed for the inclusion of temporal changes in students' transfer experiences and personal circumstances, opening up a dynamic view of mobility and related phenomenon. In this section, we highlight results for student mobility measures (both yearly and cumulative) and each of the time-varying predictors before proceeding with a discussion of limitations and implications.

Cumulative and Yearly Student Mobility Effects

The findings indicate that changing schools during the elementary school years predicts declines in classroom participation and academic performance, even when other potentially confounding predictors are included in the models. Results also show that the effect of changing schools operates a bit differently for each of these outcomes. For classroom participation, the total number of transfers a child made across the 4 years predicted lower levels of classroom participation at fifth grade. Changing schools does not, however, appear to have an immediate or year-to-year impact on classroom participation. These results are consistent with a number of mobility studies that showed that a single move may not impact children, but when moves accrue, their cumulative impact can be highly significant (Kerbow, 1996; Simpson & Fowler, 1994; Tucker et al., 1998; Wood et al., 1993). The findings for classroom participation are noteworthy, as research reporting a cumulative effect of mobility has focused on measures of academic achievement and behavior problems.

The effect of school transfers is robust and remains significant when other factors associated with school transfers are included in the equation. These factors—initial antisocial behavior, being male, initial low-income status, initial shy/withdrawn behavior, and total family stress—were also significant and appear to be strong predictors of classroom participation. Thus, the findings show that making a school change is only one of the significant factors that have a negative influence on the development of classroom participation.

The findings indicate that changing schools appears to have a different type of influence on the second significant outcome, academic performance, than it does on classroom participation. Annual and not cumulative mobility appears to affect teacher ratings of academic performance. During each school year, if a child changed schools at least once, the move or moves forecasted a decrease from the sample average slope for teacher-rated academic performance in the spring survey. Because we used a dichotomous measure, these results do not account for the multiple moves that 33 of the subjects made in a school year. We plan to examine the impact of these multiple moves in a future study.

The academic results reported here add evidence to the existing empirical literature on the association between annual school mobility and declines in academic achievement (Kerbow, 1996; Tucker et al., 1998). We were aware that a possible confounding variable might be that the children who move most often are simply the ones who would ordinarily show a decline in performance over time. However, in the final model run with the yearly transfer covariate, the total number of transfers was included on the slope. Thus, the effect was significant even when accounting for the performance of frequent movers.

We found no evidence that school changes, either the total number of transfers or the year-to-year measure of transfers, affected the growth trajectories of children's positive attitude toward school. This may be because school remains a positive influence for those who move, especially those who move often and experience other family stressors (Hetherington, 1979). The total number of transfers did play a minor but significant role in estimating the impact of teacher support on children's positive attitude toward school. Specifically, in the years that children in the sample felt significantly more teacher support than their average, their attitude scores were predicted to increase. In addition, this relationship between teacher support and attitude was stronger for children who moved more frequently. In other words, results suggest that the degree of teacher support is more important for the children who experience a school change. These findings are congruent with the research of DuBois et al. (1994), who found that young adolescents who experienced multiple economic disadvantages appeared to benefit more from social support received from adults in the school than do their "low-risk" peers.

Family Stress, Peer Acceptance, and Teacher Support Effects

In addition to examining the effects of student mobility, we considered alternative explanations for declines and increases in children's growth on the three outcomes. To this end, time-varying risk (yearly family stress) and protective factors (yearly measures of teacher support and peer acceptance) were added to hierarchical models to examine the relative explanatory strength of these factors. Of particular interest was the influence of family stress experiences because prior researchers have warned of a confounding effect of stress on school transfers. Contrary to expectations, a child's yearly experiences with family stress did not have a significant effect on any of the school engagement outcomes measured in this study.

One would expect that a child's exposure to the stress generated by divorce, job loss, or family death might result in at least temporary declines in school adjustment (DuBois et al., 1994). However, as the work of Hetherington (1979) shows, there is wide variability in the responses of children to divorce based on such factors as age, temperament, and changes in economic status. Further, not all family losses lead to greater conflict or strife within the family unit. In some cases, a parental breakup or death in the family may offer relief for children who have been living in a state of conflict or worry. Finally, a recent study by Gershoff, Aber, Raver, and Lennon (2007) suggests a much more complex relationship between family stress, low-income status, and child outcomes than the one examined in this study. Their work revealed that material hardship (i.e., months of financial troubles, food insecurity, inadequacy of medical care, residential instability) had a stronger association with family stress than income level. To more accurately assess the effect of stress, future studies of mobility should take into account these more nuanced measures of family stress and hardship.

In contrast to yearly family stress, the findings for both yearly protective factors were significant, providing evidence that social ties present in the school environment, or what some might call school social capital, serve to support successful outcomes. The findings lend support to the theory of appropriation (Vygotsky, 1962, 1978), which emphasizes that a child's adaptation to a learning environment is achieved by social means, through the people surrounding the child at school. First, yearly measures of a child's peer acceptance emerged as significant for both academic and attitudinal outcomes. In a given year, if peer acceptance was rated significantly above the child's average, his or her academic performance scores increased. One possible bias for this finding is that teacher ratings of academic performance at this age may be more a measure of how well children are getting along in class rather than a measure of achievement in math or reading. These findings should be interpreted with caution given the moderate correlation between teacher-rated peer acceptance and teacher-rated academic performance (r = .36). However, if replicated in future studies using other measures of academic growth, this finding would lend support to those who postulate that positive social interactions in the classroom facilitate intellectual development (Ladd et al., 1999).

Second, in the model testing the influence of peer acceptance on attitude toward school, positive attitudes were estimated to improve when children experienced above-average peer acceptance, though, when comparing their influence, teacher support appears to overshadow peer acceptance. These findings are consistent with developmental research showing that at this age, prior to the shift in adolescent social priorities toward peer relationships, teacher relationships are critical in terms of children's feelings about school (Birch & Ladd, 1997; Howes & Hamilton, 1993). One limitation of this study is that we did not collect data on children prior to their entry into school, during kindergarten, or during the transition to first grade. Research findings suggest that early school experiences, particularly relationships with teachers and peers, may have lasting effects on children (Birch & Ladd, 1997; Ladd et al., 1999; Pianta, Steinberg, & Rollins, 1995).

Results for the other protective factor indicate that children's perceptions of teacher support had a positive influence on the growth trajectories for children's classroom participation and positive attitudes toward school. Specifically, in a given year, when children indicated that they felt supported by their teachers significantly above their average, their classroom participation improved. In the same way, scores for positive attitude toward school improved during years when children reported more support, and, as mentioned earlier, mobile children received an additional boost in their positive feelings about school when they felt supported by teachers. The results here are not surprising given that the items we used to measure teacher support reflect a child's perspective on the condition of the relationship, emphasizing the level of helpfulness and caring present in the relationship (e.g., Does your teacher notice when you try hard? Do you like your teacher?). Prior research has established that children respond with improved attitudes and better behavior when they have closer relationships with their teacher (Birch & Ladd, 1997; Valenzuela, 1999; Wentzel, 1997). Contrary to expectation, the child's view of teacher support did not appear to impact teacher ratings of academic performance. Given that other researchers have found that warm and caring relationships with students influence academic outcomes, these results come as a surprise.

Implications for Practice in the Schools

Researchers have indicated that schools play a critical role in mitigating the negative effects of mobility on student outcomes and suggest that school personnel should be doing more to support students, given the high level of risk associated with school transfers (Mantzicopoulos & Knutson, 2000; Nelson et al., 1996; Vernberg, 1990). Unfortunately, a common concern raised by researchers is that few teachers have prior training in working with transfer students (Adelman & Taylor, 1991; Kerbow, 1996; Lash & Kirkpatrick, 1990). For example, Lash and Kirkpatrick (1990) found that teachers are trained to develop their instructional plans based on a classroom of students who will remain with them for an entire school cycle rather than assessing and teaching to the needs of students who enter in the middle of the year.

To improve outcomes for mobile students, a variety of practices have been suggested, including: intensive tutoring for high-risk transfers (Jason et al., 1992), careful assessment of social and academic strengths upon entry (Adelman & Taylor, 1991), facilitation of friendship formation and the acceptance of newcomers (Vernberg, 1990), and constructing a sense of stability and predictability in the school setting (Mantzicopoulos & Knutson, 2000). The intervention studies of high-risk elementary-age transfer students, conducted by Jason and colleagues in Chicago, have undergone the most systematic evaluation. In several controlled studies, transfer students participated in an orientation program and received biweekly tutoring with results indicating that the intervention improved student grades and achievement tests scores over the course of 1 year (Jason et al.; Jason et al., 1989, 1993). Despite including an orientation component, the intervention did not appear to positively influence social or behavioral outcomes for transfer students.

Taken together, the academic findings from the Jason studies and the behavioral and attitudinal findings of the present study hold some promise for practitioners hoping to mitigate the problems some students encounter following a school transfer. Results suggest that effective interventions for transfer students should include at least two components: (a) intensive tutoring to address the academic deficits of many transfer students and (b) university and school-based trainings to raise awareness of hardships encountered by mobile students, to increase caring responses, and to address peer acceptance in the classroom. Indeed, our previous research on school interventions indicates that a combined package of teacher, parent, and student training to support academic achievement and reduce problem behavior has impacted both school commitment and academic performance (Catalano et al., 2003).

Conclusions

One of the aims of this study was to use the developmental science framework (Magnusson & Cairns, 1996) to improve on prior research. This approach allowed us to build on the existing literature in a number of important ways. First, longitudinal data on elementary children have rarely been used in research on school mobility, especially data from studies that retained a large number of subjects. In the present investigation, all the analyses were calculated with a minimum of 94% of the original sample of preadolescent children. Even children who moved four or five times during the study years were retained in the analysis. Second, of the small group of published longitudinal studies on mobility, none have explored the relationship between school changes and classroom participation and positive attitude toward school. The findings presented in this study contribute to the existing research on the relationships between mobility, academic achievement, and behavior problems. Finally, this study demonstrated that growth curve modeling offers a powerful tool in the effort to understand how fluctuating circumstances can make a difference, both positively and negatively, for mobile children. Within the models examined in this study, a specific effort was made to shed light on changing relationships in the school that might protect children from undesirable outcomes associated with school changes.

Although the work presented here represents some key refinements in studying the developmental changes associated with mobility, a number of improvements must still be addressed. For example, within the developmental science framework, researchers are cautioned to be careful about the assumptions we make about stable factors. We chose to use a single measure of two child-level variables, antisocial behavior and low-income status, that emerged as significant predictors in most of the final models. Given their significance, future research that treats income status and antisocial behavior as time-varying factors might find that fluctuations in these variables significantly influence child outcomes. For example, a mobile child with a history of behavior problems could be influenced by a supportive teacher to make changes in classroom behavior leading to significant improvements in attitude or participation.

In outlining the developmental science framework, Magnusson and Cairns (1996) urge researchers to recognize the complexity of development and yet they acknowledge that "decisions must be made to bring some features to the foreground and move others into the background" (p. 8) in a given research project. In the present study, we used a predominantly Caucasian, suburban sample and chose to focus on protective relationships that might influence students in individual school settings. Clearly, these choices limit our ability to generalize to urban, more diverse, and more mobile populations. There is evidence in the literature that school-level effects, such as the pace of instruction and school climate, in urban schools with highly transient populations ought to be included in a longitudinal analysis of mobility (Alexander et al., 1996; Kerbow, 1996; Rumberger & Larson, 1998). Our choices also limited our ability to examine extrafamilial or neighborhood factors, including relationships with coaches, pastors, or caring neighbors who may support children new to neighborhoods and help them engage with school. Further, we did not examine the more traditional measures of social capital, including indicators of parent-community ties that may serve to support mobile students (Hagan et al., 1996; Pribesh & Downey, 1999).

In light of the limitations and strengths of this study, a promising direction for future research is to explore, using growth curve modeling, the changing family circumstances which may influence children's experience of moving. A study by Macmillan, McMorris, and Kruttschnitt (2004) marks a path in this direction by finding that the lives of young mothers and their children (ages 4–7) are bound together in significant ways. Among other things, the study shows that changes in maternal circumstances (e.g., meaningful escapes from or long-term movement into poverty) influenced children's antisocial behavior. The results are a clear invitation to include time-varying variables within the family context, such as fluctuations in family income and timing of divorce and remarriage, in future examinations of the consequences of mobility for children. Within the developmental science framework, growth curve modeling would allow for an examination of the theory (Lash & Kirkpatrick, 1990) that the reason why a family moves (e.g., upward mobility, forced mobility, lifestyle choice) leads to the great variation in consequences for children.

References

Achenbach, T. M. (1991). *Manual for the Teacher's Report Form and 1991 profile*. Burlington: University of Vermont Department of Psychiatry.

Adelman, H. S., & Taylor, L. (1991). Early school adjustment problems: Some perspectives and a project report. *American Journal of Orthopsychiatry, 61*, 468–474.

Alexander, K. L., Entwisle, D. R., & Dauber, S. L. (1993). First-grade classroom behavior: Its short- and long-term consequences for school performance. *Child Development, 64*, 801–814.

Alexander, K. L., Entwisle, D. R., & Dauber, S. L. (1996). Children in motion: School transfers and elementary school performance. *Journal of Educational Research, 90*, 3–12.

Alexander, K. L., Entwisle, D. R., & Horsey, C. S. (1997). From first grade forward: Early foundations of high school dropout. *Sociology of Education, 70*, 87–107.

Alspaugh, J. W. (1998). Achievement loss associated with the transition to middle school and high school. *Journal of Educational Research, 92*, 20–25.

Astone, N. M., & McLanahan, S. S. (1994). Family structure, residential mobility, and school dropout: A research note. *Demography, 31*, 575–584.

Birch, S. H., & Ladd, G. W. (1997). The teacher-child relationship and children's early school adjustment. *Journal of School Psychology, 35*, 61–79.

Blane, D., Pilling, D., & Fogelman, K. (1985). The use of longitudinal data in a study of children's school mobility and attainment. *British Journal of Educational Psychology, 55*, 310–313.

Brown, E. C., Catalano, R. F., Fleming, C. B., Haggerty, K. P., & Abbott, R. D. (2005). Adolescent substance use outcomes in the Raising Healthy Children Project: A two-part latent growth curve analysis. *Journal of Consulting and Clinical Psychology, 73*, 699–710.

Catalano, R. F., Mazza, J. J., Harachi, T. W., Abbott, R. D., Haggerty, K. P., & Fleming, C. B. (2003). Raising healthy children through enhancing social development in elementary school: Results after 1.5 years. *Journal of School Psychology, 41*, 143–164.

Catalano, R. F., Park, J., Harachi, T. W., Haggerty, K. P., Abbott, R. D., & Hawkins, J. D. (2005). Mediating the effects of poverty, gender, individual characteristics, and external constraints on antisocial behavior: A test of the social development model and implications for developmental life-course theory. In D. P. Farrington (Ed.), *Advances in criminological theory: Vol. 14. Integrated developmental and life-course theories of offending* (pp. 93–123). New Brunswick, NJ: Transaction.

Coie, J. D., Dodge, K. A., & Kupersmidt, J. B. (1990). Peer group behavior and social status. In S. R. Asher & J. D. Coie (Eds.), *Peer rejection in childhood* (pp. 17–59). New York: Cambridge University Press.

Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology, 94*, 95–120.

Coleman, J. S. (1990). *Foundations of social theory.* Cambridge, MA: Harvard University Press.

Cramer, W., & Dorsey, S. (1970). Are movers losers? *Elementary School Journal, 70*, 387–390.

DuBois, D. L., Felner, R. D., Meares, H., & Krier, M. (1994). Prospective investigation of the effects of socioeconomic disadvantage, life stress, and social support on early adolescent adjustment. *Journal of Abnormal Psychology, 103*, 511–522.

Durlauf, S. N. (1999). The case "against" social capital. *Focus, 20*, 1–5.

Eccles, J. S., & Midgley, C. (1990). Changes in academic motivation and self-perception during early adolescence. In R. Montemayor, G. R. Adams, & T. P. Gullotta (Eds.), *From childhood to adolescence: A transitional period (Advances in adolescent development).* (pp. 134–155). Newbury Park, CA: Sage.

Eckenrode, J., Rowe, E., Laird, M., & Brathwaite, J. (1995). Mobility as a mediator of the effects of child maltreatment on academic performance. *Child Development, 66*, 1130–1142.

Ellickson, P. L., & McGuigan, K. A. (2000). Early predictors of adolescent violence. *American Journal of Public Health, 90*, 566–572.

Felner, R. D., Ginter, M., & Primavera, J. (2002). Primary prevention during school transitions: Social support and environmental structure. In T. A. Revenson, A. R. D'Augelli, S. E. French, D. L. Hughes, D. Livert, E. Seidman, et al. (Eds), *A quarter century of community psychology: Readings from the American Journal of Community Psychology* (pp. 147–161). New York: Kluwer Academic/Plenum.

Finn, J. D. (1989). Withdrawing from school. *Review of Educational Research, 59*, 117–142.

Finn, J. D., & Rock, D. A. (1997). Academic success among students at risk for school failure. *Journal of Applied Psychology, 82*, 221–234.

Fleming, C. B., Harachi, T. W., Catalano, R. F., Haggerty, K. P., & Abbott, R. D. (2001). Assessing the effects of a school-based intervention on unscheduled school transfers during elementary school. *Evaluation Review, 25*, 655–679.

Garmezy, N. (1983). Stressors of childhood. In N. Garmezy & M. Rutter (Eds.), *Stress, coping, and development in children* (pp. 43–84). Baltimore: Johns Hopkins University Press.

Gershoff, E. T., Aber, J. L., Raver, C. C., & Lennon, M. C. (2007). Income is not enough: Incorporating material hardship into models of income associations with parenting and child development. *Child Development, 78*, 70–95.

Goldstein, L. S. (1999). The relational zone: The role of caring relationships in the co-construction of mind. *American Educational Research Journal, 36*, 647–673.

Hagan, J., MacMillan, R., & Wheaton, B. (1996). New kid in town: Social capital and the life course effects of family migration on children. *American Sociological Review, 61*, 368–385.

Haggerty, K. P., Catalano, R. F., Harachi, T. W., & Abbott, R. D. (1998). Description de l'implementation d'un programme de prévention des problèmes de comportement à l'adolescence [Preventing adolescent problem behaviors: A comprehensive intervention description]. *Criminologie, 31*, 25–47.

Haveman, R., Wolfe, B., & Spaulding, J. (1991). Childhood events and circumstances influencing high school completion. *Demography, 28*, 133–157.

Hawkins, J. D., & Catalano, R. F. (1990). *Seattle Social Development Project.* Unpublished manuscript, Social Development Research Group, University of Washington.

Heinlein, L. M., & Shinn, M. (2000). School mobility and student achievement in an urban setting. *Psychology in the Schools, 37*, 349–357.

Hendershott, A. B. (1989). Residential mobility, social support and adolescent self-concept. *Adolescence, 24*, 217–232.

Hetherington, E. (1979). Divorce: A child's perspective. *American Psychologist, 34*, 851–858.

Hirsch, B. J., & DuBois, D. L. (1992). The relation of peer social support and psychological symptomatology during the transition to junior high school: A two-year longitudinal analysis. *American Journal of Community Psychology, 20*, 333–347.

Hofferth, S. L., Boisjoly, J., & Duncan, G. J. (1998). Parents' extrafamilial resources and children's school attainment. *Sociology of Education, 71*, 246–268.

Howes, C., & Hamilton, C. E. (1993). The changing experience of child care: Changes in teachers and in teacher–child relationships and children's social competence with peers. *Early Childhood Research Quarterly, 8*, 15–32.

Hughes, J. N., Cavell, T. A., & Willson, V. (2001). Further support for the developmental significance of the quality of the teacher-student relationship. *Journal of School Psychology, 39*, 289–301.

Ingersoll, G. M., Scamman, J. P., & Eckerling, W. D. (1989). Geographic mobility and student achievement in an urban setting. *Educational Evaluation and Policy Analysis, 11*, 143–149.

Isakson, K., & Jarvis, P. (1999). The adjustment of adolescents during the transition into high school: A short term longitudinal study. *Journal of Youth and Adolescence*, *28*, 1–26.

Jason, L. A., Betts, D., Johnson, J., Smith, S., Krueckeberg, S., & Cradock, M. (1989). An evaluation of an orientation plus tutoring school-based prevention program. *Professional School Psychology*, *4*, 273–284.

Jason, L. A., Danner, K. E., Kurasaki, K. S., Halpert, J., Weine, A. M., Warren-Sohlberg, L., et al. (1993). A 1-year follow-up of a preventive program for high-risk transfer children. *Journal of Emotional and Behavioral Disorders*, *1*, 215–221.

Jason, L. A., Weine, A. M., Johnson, J. H., Warren-Sohlberg, L., Filippelli, L. A., Turner, E., et al. (1992). *Helping transfer students: Strategies for educational and social readjustment*. San Francisco: Jossey-Bass.

Kerbow, D. (1996). Patterns of urban student mobility and local school reform. *Journal of Education for Students Placed at Risk*, *1*, 147–169.

Kusche, C. A., & Greenberg, M. T. (1988). *Seattle Personality Questionnaire for Young School Aged Children*. Unpublished personality questionnaire, University of Washington, Department of Psychology.

Ladd, G. W., Birch, S. H., & Buhs, E. S. (1999). Children's social and scholastic lives in kindergarten: Related spheres of influence? *Child Development*, *70*, 1373–1400.

Ladd, G. W., & Burgess, K. B. (2001). Do relational risks and protective factors moderate the linkages between childhood aggression and early psychological and school adjustment? *Child Development*, *72*, 1579–1601.

Lash, A. A., & Kirkpatrick, S. L. (1990). A classroom perspective on student mobility. *Elementary School Journal*, *91*, 177–191.

Lord, S. E., Eccles, J. S., & McCarthy, K. A. (1994). Surviving the junior high school transition: Family processes and self-perceptions as protective and risk factors. *Journal of Early Adolescence*, *14*, 162–199.

Macmillan, R., McMorris, B. J., & Kruttschnitt, C. (2004). Linked lives: Stability and change in maternal circumstances and trajectories of antisocial behavior in children. *Child Development*, *75*, 205–220.

Magnusson, D., & Cairns, R. B. (1996). Developmental science: Toward a unified framework. In R. B. Cairns, G. H. Elder Jr., & E. J. Costello (Eds.), *Developmental science* (pp. 7–30). New York: Cambridge University Press.

Mantzicopoulos, P., & Knutson, D. J. (2000). Head Start children: School mobility and achievement in the early grades. *Journal of Educational Research*, *93*, 305–311.

Moore, K. A., Vandivere, S., & Ehrle, J. (2000). *Turbulence and child well-being* (Policy Brief No. Series B, No. B-16, the Urban Institute, National Survey of America's Families). Washington, DC: Child Trends.

Morrow, V. (1999). Conceptualising social capital in relation to the well-being of children and young people: A critical review. *Sociological Review*, *47*, 744–765.

Nelson, P. S., Simoni, J. M., & Adelman, H. S. (1996). Mobility and school functioning in the early grades. *Journal of Educational Research*, *89*, 365–369.

Pianta, R. C., Steinberg, M. S., & Rollins, K. B. (1995). The first two years of school: Teacher-child relationships and deflections in children's classroom adjustment. *Development and Psychopathology*, *7*, 295–312.

Pribesh, S., & Downey, D. B. (1999). Why are residential and school moves associated with poor school performance? *Demography*, *36*, 521–534.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Newbury Park, CA: Sage.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., & Congdon, R. T. (2000). *HLM 5: Hierarchical linear and nonlinear modeling (Version 5)*. Lincolnwood, IL: Scientific Software International.

Reynolds, A. J., & Bezruczko, N. (1993). School adjustment of children at risk through fourth grade. *Merrill-Palmer Quarterly*, *39*, 457–480.

Rumberger, R. W., & Larson, K. A. (1998). Student mobility and the increased risk of high school dropout. *American Journal of Education*, *107*, 1–35.

Rumberger, R. W., Larson, K. A., Ream, R. K., & Palardy, G. J. (1999). *The educational consequences of mobility for California students and schools* (PACE Policy Brief, Vol. 1, No. 1). Berkeley: University of California.

Sandler, I. N., Miller, P., Short, J., & Wolchik, S. A. (1989). Social support as a protective factor for children in stress. In D. Belle (Ed.), *Children's social networks and social supports* (pp. 277–307). Oxford, England: Wiley.

Schaller, J. (1976). Geographic mobility as a variable in ex-post facto research. *British Journal of Educational Psychology*, *46*, 341–343.

Seidman, E., & French, S. E. (1997). Normative school transitions among urban adolescents: When, where, and how to intervene. In H. J. Walberg, O. Reyes, & R. P. Weissberg (Eds.), *Children and youth: Interdisciplinary perspectives* (pp. 66–189). Thousand Oaks, CA: Sage.

Simmons, R. G., Burgeson, R., Carlton-Ford, S., & Blyth, D. A. (1987). The impact of cumulative change in early adolescence. *Child Development*, *58*, 1220–1234.

Simpson, G. A., & Fowler, M. G. (1994). Geographic mobility and children's emotional/behavioral adjustment and school functioning. *Pediatrics*, *93*, 303–309.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Thousand Oaks, CA: Sage.

South, S. J., Haynie, D. L., & Bose, S. (2007). Student mobility and school dropout. *Social Science Research*, *36*, 68–94.

Temple, J. A., & Reynolds, A. J. (1999). School mobility and achievement: Longitudinal findings from an urban cohort. *Journal of School Psychology*, *37*, 355–377.

Texas Department of Education. (1997). *A study of student mobility in the Texas Public Schools*. Report No. 3, Statewide Texas Educational Progress Study.

Thompson, R. A. (1998). Early sociopersonality development. In W. Damon & N. Eisenberg (Eds.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (5th ed., pp. 25–86). New York: Wiley.

Tucker, C., Marx, J., & Long, L. (1998). "Moving on": Residential mobility and children's school lives. *Sociology of Education, 71,* 111–129.

Valenzuela, A. (1999). *Subtractive schooling: U.S.-Mexican youth and the politics of caring.* Albany: State University of New York Press.

van der Veer, R., & Valsiner, J. (Eds.). (1994). *The Vygotsky reader.* Cambridge, MA: Blackwell.

Vernberg, E. M. (1990). Experiences with peers following relocation during early adolescence. *American Journal of Orthopsychiatry, 60,* 466–472.

Vygotsky, L. S. (1962). *Thought and language.* Cambridge, MA: MIT Press.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Cambridge, MA: Harvard University Press.

Walker, H. M., & McConnell, S. R. (1988). *The Walker-McConnell Scale of Social Competence and School Adjustment.* Austin, TX: Pro-Ed.

Wentzel, K. R. (1993). Does being good make the grade? Social behavior and academic competence in middle school. *Journal of Educational Psychology, 85,* 357–364.

Wentzel, K. R. (1997). Student motivation in middle school: The role of perceived pedagogical caring. *Journal of Educational Psychology, 89,* 411–419.

Wenz-Gross, M., Siperstein, G. N., Untch, A. S., & Widaman, K. F. (1997). Stress, social support, and adjustment of adolescents in middle school. *Journal of Early Adolescence, 17,* 129–151.

Werthamer-Larsson, L., Kellam, S. G., & Ovesen-McGregor, K. E. (1990). Teacher interview: Teacher Observation of Classroom Adaptation–Revised (TOCA–R). In S. G. Kellam (Ed.), *Johns Hopkins Prevention Center training manual.* Baltimore: Johns Hopkins University.

Wertsch, J. V. (1985). *Vygotsky and the social formation of the mind.* Cambridge, MA: Harvard University Press.

Wood, D., Halfon, N., Scarlata, D., Newacheck, P., & Nessim, S. (1993). Impact of family relocation on children's growth, development, school function, and behavior. *Journal of the American Medical Association, 270,* 1334–1338.

Child Development, November/December 2008, Volume 79, Number 6, Pages 1853 – 1868

Early Histories of School-Aged Children With Attention-Deficit/Hyperactivity Disorder

Irene M. Loe
Stanford University School of Medicine

Maria D. Balestrino
Lake Erie College of Osteopathic Medicine

Randall A. Phelps
Oregon Health and Science University

Marcia Kurs-Lasky
University of Pittsburgh

Diego Chaves-Gnecco and Jack L. Paradise
University of Pittsburgh School of Medicine and Children's Hospital of Pittsburgh

Heidi M. Feldman
Stanford University School of Medicine

In a prospective study of developmental outcomes in relation to early-life otitis media, behavioral, cognitive, and language measures were administered to a large, diverse sample of children at 2, 3, 4, 6, and 9 – 11 years of age (N = 741). At 9 – 11 years of age, 9% of the children were categorized as having attention-deficit/hyperactivity disorder (ADHD) based on parent report. Compared to the non-ADHD group, the ADHD group had higher (i.e., less favorable) scores on parent and teacher versions of the Child Behavior Checklist at all ages. Children in the ADHD group also had lower scores on cognitive and receptive language measures in preschool. The findings support the concept that ADHD is a cognitive as well as a behavioral disorder.

The diagnosis of attention-deficit/hyperactivity disorder (ADHD) implies long-standing behavioral problems involving either inattention or hyperactivity – impulsivity or both (American Psychiatric Association, 1994). According to statistics from the 2003 National Health Interview Survey, the prevalence in that year of ADHD for U.S. children 3 – 17 years of age was 6% (Dey & Bloom, 2005), although a summary of epidemiologic studies in children 6 – 12 years of age reported a wider range of 4% – 12% (Brown et al., 2001). Overall, the prevalence of ADHD is at least 4 times higher in boys than in girls (Biederman et al., 2002; Gaub & Carlson, 1997). The wide reported range in prevalence may in part reflect the heterogeneity of the populations studied. Thus, the prevalence of ADHD is increased in certain populations, such as children born preterm (Bhutta, Cleves, Casey, Cradock, & Anand, 2002; Botting, Powls, Cooke, & Marlow, 1997) or who had low birth weights (Gray, Indurkhya, & McCormick, 2004; McCormick, Workman-Daniels, & Brooks-Gunn, 1996). Children of adolescent mothers are also at an increased risk of developing behavioral problems at school, hyperactivity, and learning disabilities (Camp, 1996). The prevalence of ADHD in children free of these biological and social risk factors has not been reported.

ADHD is associated with a wide range of functional impairments, including academic problems, educational underachievement, and peer social problems (Barkley, 2006; Weiss & Hechtman, 1993). The poor long-term school outcomes, including increased school absenteeism, grade retention, and dropout, of children with ADHD have been well documented in both population-based studies (Barbaresi, Katusic, Colligan, Weaver, & Jacobsen, 2007) and clinic-referred populations (for a review, see Loe & Feldman, 2007).

Typically, the diagnosis of ADHD is established when children are in elementary school. Several longitudinal studies have documented the persistence of behavioral characteristics of the disorder into adolescence and adulthood (Barkley, 2006; Biederman et al., 1998; Mannuzza, Klein, & Moulton, 2003). Many of the early longitudinal studies of children with ADHD focused mainly on boys from clinic-referred samples

---

This work was supported by Grants HD26026 and HD42080 from the National Institute of Child Health and Human Development and the Agency for Healthcare Research and Quality and by Grants T73 MC00036 and T77 MC00031 from the Maternal and Child Health Bureau of the Health Resources and Services Administration of the U.S. Department of Health and Human Services. We thank Lynne Huffman, MD, for her helpful statistical support and review of the article.

Correspondence concerning this article should be addressed to Heidi M. Feldman, Department of Pediatrics, Stanford University School of Medicine, 750 Welch Road, Suite 315, Palo Alto, CA 94304. Electronic mail may be sent to hfeldman@stanford.edu.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7906-0022

who were recruited at school age (Mannuzza, Klein, Bonagura, & Konig, 1988; Weiss & Hechtman, 1993). Another limitation of the studies was the recruitment of matched controls at the time of follow-up rather than at the start of the prospective study. Despite these limitations, the consensus is that symptoms of ADHD persist after childhood. Approximately 65% of children with ADHD continue to meet diagnostic criteria for the condition as adolescents (August, Braswell, & Thuras, 1998; Ingram, Hechtman, & Morgenstern, 1999). Conversely, adolescents who are diagnosed with behavioral disorders commonly have a history of behavioral problems during the preschool years (Campbell, Shaw, & Gilliom, 2000; Loeber & Dishion, 1983).

Many cross-sectional studies rely on retrospective reports and have limited or no objective measures of behavior and cognition during the toddler and preschool years. In the absence of prospective, longitudinal studies, it is comparatively more difficult to learn about the early childhood behavioral or cognitive characteristics of children who are ultimately diagnosed with the disorder. Nonetheless, the developmental sequence of symptoms is important for understanding the origins of ADHD. If ADHD is primarily the result of a poor fit between active, exuberant children and demanding school environments, then children diagnosed at school age as having ADHD might have been indistinguishable from non-ADHD peers during their preschool years. In contrast, if the diagnosis of ADHD represents the culmination of long-standing behavioral and cognitive differences, then children diagnosed at school age as having ADHD should have had more behavioral problems and cognitive weaknesses than peers during their preschool years. If symptoms or precursors of ADHD can be identified during the early preschool years, such findings might help in the early identification of children with, or at risk for, the disorder and in turn might point to interventions aimed at improving long-term academic, educational, and social outcomes.

Several lines of evidence suggest that the behavioral characteristics of ADHD emerge during the preschool period. First, adolescents who are diagnosed with behavioral disorders commonly have a history of behavioral problems during the preschool period (Farrington, Loeber, & Van Kammen, 1990; Pardini, Obradovic, & Loeber, 2006). An obvious limitation of such research, however, is that the validity of the retrospective reports is inherently uncertain. Second, a cross-sectional study involving large samples of preschool, elementary, and secondary school children found teacher-reported symptoms of ADHD in 18.2% of the preschool children and slightly lower proportions among older children (Nolan, Gadow, & Sprafkin, 2001). Such prevalence data do not clarify whether the children with early symptoms would be the same ones who would later be diagnosed with ADHD.

Longitudinal studies that enrolled children as preschoolers also support the concept that symptoms of ADHD emerge in the preschool period. One such study recruited children who had been diagnosed with ADHD at 4–6 years of age from psychiatric clinics. These children tended to show persistence of the disorder as they entered school age (Lahey, Pelham, Loney, Lee, & Willcutt, 2005) though the subtype of ADHD sometimes shifted as children grew older; hyperactivity symptoms tended to decline from preschool to school age, whereas inattention symptoms remained stable. A limitation of such an approach is that it involved a clinical rather than a community sample. Children diagnosed at early ages may be among the most severely affected with the disorder, accounting for both the early age of identification and the continuity of symptoms. In a different experimental design, a longitudinal study of a small sample of twins discordant for attentional problems and ADHD at school age found that the twin with ADHD had more behavioral problems and attentional weaknesses as early as 2 years of age. However, the sample size of this study was small, and the circumstances of twinning are distinctive. Moreover, the affected twins also had delayed motor and slowed physical growth, suggesting that the affected child had more pervasive vulnerabilities (Lehn et al., 2007).

Campbell and colleagues conducted a longitudinal study of boys identified through flyers in preschool programs and pediatric offices, aimed at determining the continuity of symptoms of ADHD. The study followed "hard-to-manage" preschool boys and controls from ages 3 to 9 in the laboratory, home, and school settings (Campbell, Pierce, March, Ewing, & Szumowski, 1994; Campbell, Pierce, Moore, & Marakovitz, 1996; Campbell et al., 2000; Marakovitz & Campbell, 1998). The study excluded girls. Compared with control boys, at 4 and 6 years of age, the hard-to-manage group was more hyperactive, disruptive, irritable, and aggressive. By age 9, 36.5% of the hard-to-manage boys met the criteria for ADHD. Consistent with other studies, these results suggest that early onset of significant behavioral problems and hyperactivity is likely to be associated with persistence of those problems (Green, O'Reilly, Itchon, & Sigafoos, 2005; Sonuga-Barke, Auerbach, Campbell, Daley, & Thompson, 2005). Because the study was limited to boys with early behavioral symptoms, it is correspondingly limited in helping us understand the developmental course of children generally who are later identified as having ADHD.

Hyperactivity in preschool children is associated not only with later behavioral problems but also with mild intellectual impairment and poor preacademic skills (DuPaul, McGoey, Eckert, & VanBrakle, 2001; Sonuga-Barke et al., 2005). In the study cited previously, Marakovitz and Campbell (1998) reported that hard-to-manage boys also had significantly lower cognitive test scores at 4 and 6 years of age than the comparison group. A 12-year follow-up study of preschool children identified as "pervasively hyperactive" showed persistently poorer cognitive skills and lower levels of reading ability compared with their nonhyperactive peers (McGee, Partridge, Williams, & Silva, 1991). Correlations between externalizing behavioral problems and expressive language delays in toddlers also have been reported (Horwitz et al., 2003; Stevenson & Richman, 1978). However, the long-term outcomes of learning and academic skills in these children have yet to be elucidated.

The goal of the present study was to describe behavioral and cognitive antecedents of ADHD. This study improves on previously reported studies by considering a large and sociodemographically diverse community-based sample of boys and girls. In addition, the children were all full term, had normal birth weights, and were healthy at birth. Their mothers were all older than 18 years of age. Thus, the sample was at comparatively low risk for ADHD on the basis of these biological and social risk factors. The data were drawn from a large, longitudinal study of possible developmental consequences of early-life otitis media. The study participants had been followed prospectively from early infancy and had undergone developmental assessment at 2, 3, 4, 6, and 9–11 years of age, a longer time span than in several of the earlier studies cited previously. Results of this study have consistently shown that the duration of middle ear effusion (MEE) from otitis media, within the limits studied, is not associated with scores on a wide range of language, intellectual, academic, behavioral, and attentional measures at any of the ages tested (Paradise et al., 2000, 2003; Paradise et al., 2001; Paradise et al., 2005; Paradise et al., 2007). The present study compares the behavioral scores obtained at ages 2 to 9–11 years and the cognitive and receptive language scores obtained at ages 2–6 years, in children classified at 9–11 years of age as having ADHD, with corresponding scores in children in that age period classified as not having ADHD.

Method

Participants

The children in the present sample were enrolled in a prospective study that addressed possible developmental consequences of early-life otitis media (Paradise et al., 1997). A total of 6,350 infants less than 2 months of age were enrolled in the study at two urban hospitals, Children's Hospital of Pittsburgh and Mercy Hospital of Pittsburgh, and six nearby private pediatric group practices (two small town/rural and four suburban private practices) over a period of more than 4 years. The demographics of the sample resemble those of the greater Pittsburgh metropolitan area, consisting of primarily Caucasian and African American children and relatively few Hispanic, Asian, and no Native American children. The study's exclusion criteria ensured that the children were healthy at birth and at low risk for neurodevelopmental disorders. Infants who met any of the following criteria were excluded: birth weight less than 5 lb (2,268 g); small for gestational age; a history of neonatal asphyxia or other serious illness; a major congenital abnormality or chronic illness; multiple birth; a sibling enrolled in the study; in foster care or adopted; mother dead, seriously ill, a known drug or alcohol abuser, or, in the judgment of study personnel, too limited socially or intellectually to give informed consent or adhere to the study protocol; mother aged 18 years and younger; and English not the only household language. All children had parental consent to participate in the large study.

The study included two main components. One component was a randomized clinical trial in which children who met specified criteria regarding persistent MEE during the first 3 years of life were randomly assigned to have tympanostomy tubes inserted either promptly or up to 9 months later if MEE persisted. The other component consisted of a representative subgroup of enrolled children who ranged from those with no MEE to those with durations of effusion just short of meeting the study's criteria for randomization. These children were followed in the same fashion and underwent the same schedule of developmental assessments as the randomized children as described below. The study was approved by the institutional review boards of both participating hospitals (Paradise et al., 1997; Paradise et al., 2000; Paradise et al., 2001).

For the present study, we analyzed data from the 741 enrolled children who underwent developmental evaluation at 9–11 years of age. Figure 1 describes the subject selection for the current report. Larger proportions of children eligible for the randomized

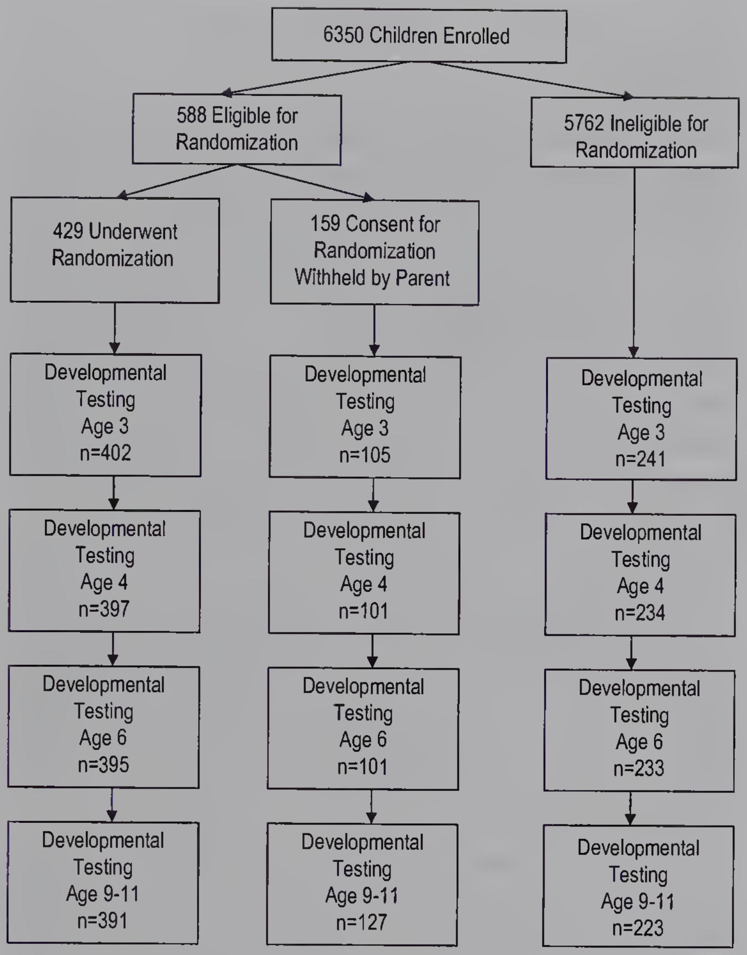

6350 Children Enrolled
588 Eligible for Randomization
5762 Ineligible for Randomization
429 Underwent Randomization
159 Consent for Randomization Withheld by Parent
Developmental Testing Age 3 n=402
Developmental Testing Age 3 n=105
Developmental Testing Age 3 n=241
Developmental Testing Age 4 n=397
Developmental Testing Age 4 n=101
Developmental Testing Age 4 n=234
Developmental Testing Age 6 n=395
Developmental Testing Age 6 n=101
Developmental Testing Age 6 n=233
Developmental Testing Age 9-11 n=391
Developmental Testing Age 9-11 n=127
Developmental Testing Age 9-11 n=223

Figure 1. Enrollment, randomization, follow-up, and developmental testing of the children.

clinical trial compared to those in the subgroup not eligible for randomization were Black, from urban areas, and from families of lower socioeconomic status (SES; based on maternal level of education and type of health insurance; Paradise et al., 1997). The children in the subgroup whose duration of otitis media with effusion did not reach the threshold for inclusion in the trial represented the range of MEE and SES of the original group of 6,350 children enrolled. The institutional review board of the University of Pittsburgh deemed the present analysis exempt from review.

Procedures

Parent interview. As a component of the visit when the children were 9–11 years old, study personnel interviewed parents about their children's health and neurodevelopmental status. Questions included (a) whether the child had any medical problems or had received any medical diagnoses, including neurobehavioral diagnoses such as ADHD; (b) whether any diagnoses not made were suspected; (c) whether the child was receiving stimulant medication or atomoxetine; (d) whether the child's parents or teacher had concerns about the child's learning abilities; and (e) whether the child was enrolled in special education or Title I reading or mathematics programs or was receiving after-school tutoring or any other school-based or private remedial service. We considered parents as valid informants of this information. Study personnel also collected demographic information, including the child's gender and race, place of residence, health insurance status, parents' educational attainment, and whether the parents lived in the home. Type of health insurance and maternal education were used as indicators of SES. Whether or not fathers were living in the home was used as a measure of environmental influence.

Criteria for classifying children. We classified children into the ADHD group if parents reported that (a) the child had been given the diagnosis of ADHD or (b) the diagnosis was suspected and the child was receiving a stimulant medication or atomoxetine. Only three parents reported that the children were suspected of having ADHD and were on medication. If parents reported either that (a) ADHD was suspected but that the child was not receiving medication or (b) ADHD was not suspected, the child was classified into the non-ADHD group.

Behavioral characteristics of participants at ages 9–11 years. We used the Child Behavior Checklist (CBCL) parent version (Achenbach, 1991a, 1992) and the CBCL teacher version or Teacher Report Form (Achenbach, 1991a) to assess behavioral characteristics of children at ages 9–11 years. These questionnaires ask informants to rate children on over 100 behavioral problems using a 3-point Likert scale, scoring statements as *not true, somewhat or sometimes true*, or *very or often true*. The results are presented as an age-adjusted *T* score (mean of 50 and standard deviation of 10) for the entire set (Total *T* score), externalizing (Delinquent Behavior, Aggressive Behavior) problems (Externalizing *T* score), and internalizing (Withdrawn, Somatic Complaints, Anxious/Depressed) problems (Internalizing *T* score). In addition, the questionnaire generates narrow-band *T* scores concerning eight different domains or subscales of behavior—Withdrawn, Somatic Complaints, Anxious/Depressed, Social Problems, Thought Problems, Attention Problems, Delinquent Behavior, and Aggressive Behavior. Higher *T* scores indicate larger numbers of behavioral problems and thus less favorable functioning. In the CBCL 4–18 versions used in this study, a Total Problems *T* score of 60–63 represented the 82nd–90th percentile for the normative sample studied; we chose this range as the borderline clinical range for the present analysis. However, the CBCL developers note that *T* scores from 60 to 63 are less clearly deviant than scores above 63 and highlight that cutoff points other than 60 may be more effective in particular samples (Achenbach, 1991b). A *T* score of 70 is considered clinically significant and represents the 97.7th percentile.

Classifying children into a learning problems group. We classified children as having learning problems if (a) the parent reported that either the parent or the teacher suspected a learning disorder and (b) the child was receiving any type of school-based, after-school, or private remedial support services. We accepted reports of any remedial educational service because the parents did not differentiate among special education services, Title I reading or mathematics support, academic accommodations, or other academic support services in their responses.

Early childhood measures. Table 1 summarizes the developmental assessments that had been performed when the children were 2, 3, 4, 6, and 9–11 years of age and that figure in the present analysis. All the children were tested individually within 2 months of their birthdays when their hearing was normal (unless they would not be able to be rescheduled within the required time frame) by trained personnel unaware of their medical history or SES. We used the CBCL parent version (Achenbach, 1991a, 1992) to assess behavioral characteristics of children each year from 2 to 6 and at 9–11 years of age and also the CBCL teacher version (Achenbach, 1991a) at 6 and 9–11 years of age.

To measure cognition when the children were 3 and 4 years of age, we used the McCarthy Scales of Children's Abilities General Cognitive Index (MSGCI; McCarthy, 1972). This test measures verbal, perceptual–performance, and quantitative skills and provides a composite score called the General Cognitive Index (GCI). Among the cognitive and intellectual measures available at the time of initiation of the study, we chose the McCarthy Scales of Children's Abilities because it was a well-respected and broad test of abilities that included verbal and nonverbal tests. The GCI has a mean of 100 and a standard deviation of 15, whereas the subtests have a mean of 50 and standard deviation of 10. To measure intelligence when the children were 6 years of age, we used the Wechsler Intelligence Scale for Children–Third Edition (WISC–III; Wechsler, 1991). The Full-Scale Intelligence Quotient (FSIQ), which has a mean of 100 and standard deviation of 15, integrates information from the verbal and performance scales. To assess receptive vocabulary at 3, 4, and 6 years of age as one aspect of language functioning, we used the Peabody Picture Vocabulary Test–Revised, Form M (PPVT–R; Dunn, 1981). The test has a mean of 100 and standard deviation of 15. Lower scores on all these measures indicate less favorable functioning.

Results

We used nondirectional tests for all analyses and set statistical significance at $p \leq .05$. There were 741 children included in the study. Of these, 69 children (9%) were classified into the ADHD group; within the ADHD group, 63 (91%) were being treated with stimulant medications or atomoxetine. Children who were only suspected of having ADHD but were not receiving medication were classified in the non-ADHD group; removing from the analysis data from these children does not substantively change the results.

Characteristics of Children According to ADHD Group

Table 2 summarizes selected demographic characteristics of the children we evaluated. We used

Table 1
Selected Assessments Performed at the Ages of 2, 3, 4, 6, and 9–11

		Age (years) at which assessment performed				
Domain	Test	2	3	4	6	9–11
Behavior	Child Behavior Checklist, parent version (2–3)	X	X			
	Child Behavior Checklist, parent version (4–18)			X	X	X
	Child Behavior Checklist, teacher version				X	X
Cognition	General Cognitive Index of McCarthy Scales of Children's Abilities		X	X		
Intelligence	Full-Scale Intelligence Quotient of Wechsler Intelligence Scale for Children–Third Edition				X	
Receptive vocabulary	Peabody Picture Vocabulary Test–Revised, Form M		X	X	X	

Table 2
Characteristics of the Children Classified as Having or Not Having ADHD

Characteristics	Study group			
	ADHD (n = 69)	Non-ADHD (n = 672)	Total (N = 741)	p value
Sex				
Male	54 (78.3)	357 (53.1)	411 (55.5)	< .001
Female	15 (21.7)	315 (46.9)	330 (44.5)	
Race				
White	46 (66.7)	436 (64.9)	482 (65.0)	.77
Non-White	23 (33.3)	236 (35.1)	259 (35.0)	
Study site grouping				
Urban	34 (49.3)	294 (43.8)	328 (44.3)	.07
Small town/rural	27 (39.1)	224 (33.3)	251 (33.9)	
Suburban	8 (11.6)	154 (22.9)	162 (21.9)	
Type of insurance				
Medicaid	35 (50.7)	218 (32.4)	253 (34.1)	.004
Private	32 (46.4)	415 (61.8)	447 (60.3)	
CHIP	0 (0.0)	24 (3.6)	24 (3.2)	
None	2 (2.9)	15 (2.2)	17 (2.3)	
Mother's level of education				
Less than high school graduate	4 (5.8)	35 (5.2)	39 (5.3)	.059
High school graduate	58 (84.1)	496 (73.8)	554 (74.8)	
College or postgraduate	6 (8.7)	140 (20.8)	146 (19.7)	
Unknown	1 (1.4)	1 (0.15)	2 (0.3)	
Father's level of education				
Less than high school graduate	12 (17.4)	40 (6.0)	52 (7.0)	.001
High school graduate	44 (63.8)	484 (72.0)	528 (71.3)	
College or postgraduate	8 (11.6)	123 (18.3)	131 (17.7)	
Unknown	5 (7.2)	25 (3.7)	30 (4.0)	
Mother in the home				
Yes	66 (95.7)	647 (96.3)	713 (96.2)	.80
No	3 (4.3)	25 (3.7)	28 (3.8)	
Father in the home				
Yes	32 (46.4)	408 (60.7)	439 (59.2)	.02
No	37 (53.6)	264 (39.3)	302 (40.8)	
Concurrent learning problems				
Yes	47 (68.1)	141 (21.0)	188 (25.4)	< .001
No	22 (31.9)	531 (79.0)	553 (74.6)	

Note. CHIP = Children's Health Insurance Program; ADHD = attention-deficit/hyperactivity disorder.

chi-square analyses to evaluate differences in proportions between the ADHD and the non-ADHD groups. Compared with the non-ADHD group, significantly larger proportions of children in the ADHD group were boys, $\chi^2(1) = 16.005$, $p < .001$; were enrolled in Medicaid insurance, $\chi^2(2) = 11.10$, $p = .004$ (excluding children with no health insurance or whose status was unknown); had parents with lower educational attainment; fathers, $\chi^2(2) = 14.2$, $p = .001$; trend for mothers, $\chi^2(2) = 5.6$, $p = .059$; and had fathers not living in the home, $\chi^2(1) = 5.3$, $p = .02$. Using a t statistic, we found that children in the ADHD group had a higher mean percent time with MEE during their first 3 years of life (M = 34.7%, SE = 1.61) than children in the non-ADHD group (M = 29.0%, SE = 0.75). This difference was significant $t(81.1) = 3.24$, $p = .002$, but was no longer significant after adjusting for child gender, type of health insurance, whether or not the father was living in the home, and level of maternal education, $F(1, 534) = 2.45$, $p = .118$.

Forty-seven of 69 children in the ADHD group (68%) also met criteria for having learning problems compared with 141 of 672 children in the non-ADHD group (21%), $\chi^2(1) = 73.4$, $p < .001$. Within each group, there were no significant interactions between any of the listed demographic variables and learning problems status.

Early Behavioral and Cognitive Measures in Children With and Without ADHD

Behavioral measures. We used repeated measures analysis of variance (ANOVA) to compare the two groups on developmental patterns of behavioral symptoms, as measured by parent CBCL ratings of children at ages 2, 3, 4, 6, and 9–11 years and teacher CBCL ratings of the children at ages 6 and 9–11 years. In these analyses, age was the within-subject variable and group (ADHD vs. non-ADHD) was the between-subject variable. Because sociodemographic variables were significantly related to group status, repeated measures analyses were conducted using specific sociodemographic variables (i.e., type of health insurance, level of maternal education, whether or not the father was living in the home) as covariates. Although paternal education also was significantly different between the groups, further examination revealed that for children whose fathers were not living in the home, data regarding paternal education were missing. Therefore, maternal education was used as the covariate.

Parent CBCL. Regarding the children's Total Problems *T* scores on the parent version of the CBCL at ages 2, 3, 4, 6, and 9–11 years, there was a main effect of group, with children in the ADHD group having consistently higher (i.e., less favorable) mean Total Problems *T* scores than children in the non-ADHD group, $F(1, 565) = 51.8$, $p < .001$. The difference between the groups was greatest at 9–11 years of age, when the ADHD group had a mean *T* score of 60.9 ± 11.1 compared with 47.1 ± 10.5 for the non-ADHD group. On repeated measures analysis, there was a significant main effect of time, $F(4, 562) = 6.36$, $p < .001$, and there was a significant Group × Time interaction, indicating divergence of the scores over time between the groups, $F(4, 562) = 11.3$, $p < .001$. Analysis of Externalizing *T* scores and Internalizing *T* scores showed similar patterns of results. After adjusting for type of health insurance, level of maternal education, and whether or not the father was living in the home, all differences between ADHD and non-ADHD groups remained as follows: (a) Total Problems *T* score: main effect of group, $F(1, 560) = 44.5$, $p < .001$, and of time, $F(4, 557) = 2.55$, $p = .039$, as well as Group × Time interaction, $F(4, 557) = 10.58$, $p < .001$; (b) Externalizing *T* score: main effect of group, $F(1, 559) = 42.56$, $p < .001$, and of time, $F(4, 556) = 4.21$, $p = .002$, as well as Group × Time interaction, $F(4, 556) = 6.32$, $p < .001$; and (c) Internalizing *T* score: main effect of group, $F(1, 560) = 18.16$, $p < .001$, and of time, $F(4, 557) = 5.24$, $p < .001$, as well as Group × Time interaction, $F(4, 557) = 4.42$, $p = .002$.

Repeated measures ANOVA was conducted to compare developmental patterns of behavioral symptoms, as measured by parent CBCL ratings of children at ages 2, 3, 4, 6, and 9–11 years using four groups: ADHD only, learning problems only, both ADHD and learning problems, and neither ADHD nor learning problems. Regarding the children's Total Problems *T* scores on the parent CBCL at ages 2, 3, 4, 6, and 9–11 years, the main effect of group remained, $F(3, 563) = 22.58$, $p < .001$. The significant main effect of time on repeated measures analysis also remained, $F(4, 560) = 6.97$, $p < .001$, and there was a significant Group × Time interaction, indicating divergence of scores over time, $F(12, 1{,}686) = 4.84$, $p < .001$. The results are summarized in Figure 2. Analysis of Externalizing *T* scores and Internalizing *T* scores showed similar patterns of results. After adjusting for covariates (type of health insurance, level of maternal education, and whether or not the father was living in the home), the differences remained as follows: (a) Total Problems *T* score: main effect of group, $F(3, 558) = 19.47$, $p < .001$, and of time, $F(4, 555) = 2.65$, $p = .033$, as well as Group × Time interaction, $F(12, 1671) = 4.46$, $p < .001$; (b) Externalizing *T* score: main effect of group, $F(3, 557) = 18.56$, $p < .001$, and of time, $F(4, 554) = 4.35$, $p = .002$, as well as Group × Time interaction, $F(12, 1668) = 2.87$, $p = .001$; and (c) Internalizing *T* score: main effect

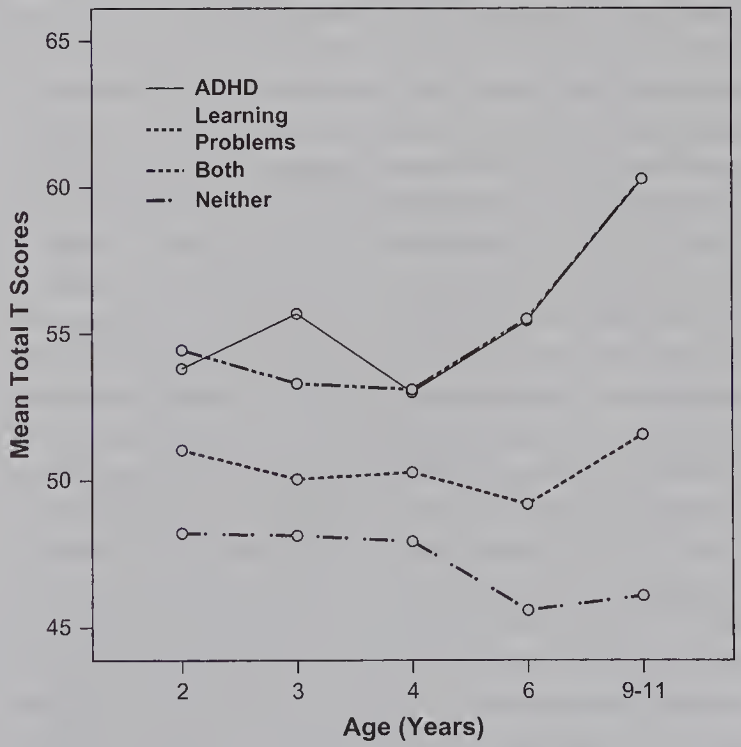


Figure 2. Mean Total Problems *T* scores on the Child Behavior Checklist, parent version, for children with ADHD only, learning problems only, both ADHD and learning problems, and neither ADHD nor learning problems at ages 2, 3, 4, 6, and 9–11 years. *Note.* ADHD = attention-deficit/hyperactivity disorder.

of group, $F(3, 558) = 12.31$, $p < .001$, and of time $F(4, 555) = 4.94$, $p = .001$, as well as Group × Time interaction, $F(12, 1671) = 4.65$, $p < .001$.

Teacher CBCL. Repeated measures ANOVA was conducted on data collected on the teacher CBCL regarding the children at ages 6 and 9–11 years. The results showed significant differences between the ADHD and the non-ADHD groups that remained significant after adjustment for type of health insurance, level of maternal education, and whether or not the father was living in the home. Compared with the non-ADHD group, the ADHD group had higher (less favorable) mean Total Problems *T* scores, $F(1, 630) = 58.83$; Externalizing *T* scores, $F(1, 630) = 30.02$; and Internalizing *T* scores, $F(1, 630) = 26.14$, all $ps < .001$. There were no significant main effects of time ($p > .05$). In addition to the main effect of group, there were significant Group × Time interactions for the Total Problems *T* score, $F(1, 630) = 4.95$, $p = .026$, and the Externalizing *T* score, $F(1, 630) = 9.48$, $p = .002$, but not for the Internalizing *T* score, $F(1, 630) = 0.25$, $p = .615$.

Repeated measures ANOVA was also conducted to compare developmental patterns of behavioral symptoms, as measured by teacher CBCL ratings of children at ages 6 and 9–11 years, again using four groups: ADHD only, learning problems only, both ADHD and learning problems, and neither ADHD nor learning problems. Regarding the children's Total Problems *T* scores on the teacher CBCL, the main effect of group remained, $F(3, 633) = 59.73$, $p < .001$. As in the two-group analysis, there was no significant main effect of time, $F(1, 633) = 2.43$, $p = .120$. In addition, there was a trend suggesting Group × Time interaction, $F(3, 633) = 2.61$, $p = .051$. Analysis of Externalizing *T* scores and Internalizing *T* scores revealed similar results. After adjusting for type of health insurance, level of maternal education, and whether or not the father was living in the home, the differences remained as follows: (a) Total Problems *T* score: main effect of group, $F(3, 628) = 52.26$, $p < .001$, no effect of time, $F(1, 628) = 0.33$, $p = .567$, as well as a trend for Group × Time interaction, $F(3, 628) = 2.57$, $p = .053$; (b) Externalizing *T* score: main effect of group, $F(3, 628) = 22.62$, $p < .001$, no effect of time, $F(1, 628) = 1.97$, $p = .161$, as well as Group × Time interaction, $F(3, 628) = 4.76$, $p = .003$; and (c) Internalizing *T* score: main effect of group, $F(3, 628) = 24.82$, $p < .001$, no effect of time, $F(1, 628) = 0.64$, $p = .424$, and a lesser trend for Group × Time interaction, $F(3, 628) = 2.26$, $p = .081$. Figure 3 summarizes the children's Total Problems *T* scores on the teacher CBCL for the four groups at ages 6 and 9–11 years. Possible interactions between ADHD and learning problems groups were investigated using multivariate ANOVA (MANOVA) and are discussed in the Results section.

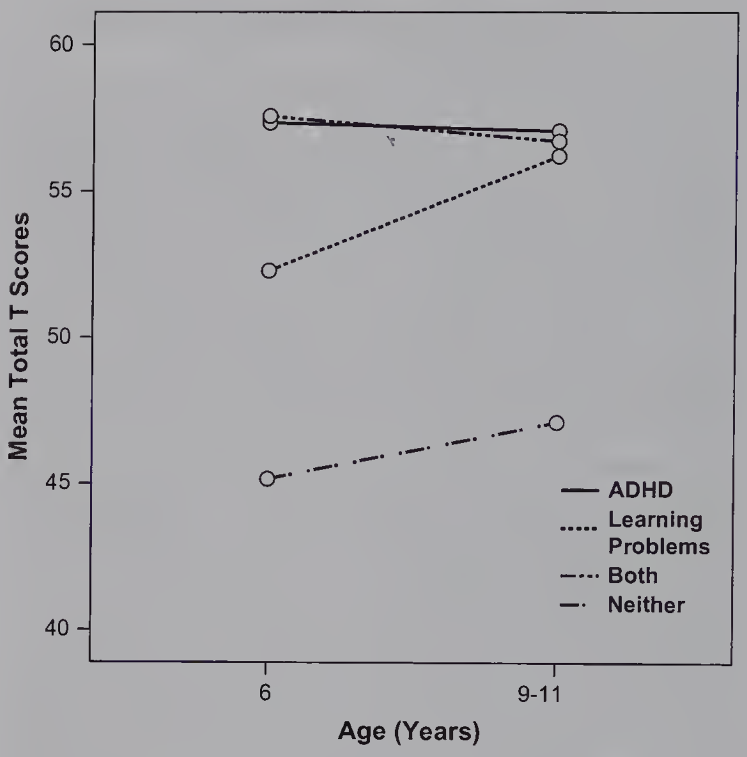


Figure 3. Mean Total Problems *T* scores on the Child Behavior Checklist, teacher version, for the children with ADHD only, learning problems only, both ADHD and learning problems, and neither ADHD nor learning problems at ages 6 and 9–11 years. *Note.* ADHD = attention-deficit/hyperactivity disorder.

Narrow-band scores on the parent and teacher CBCLs. Narrow-band scores on subscales of the parent CBCL were evaluated at ages 4, 6, and 9–11 years using the Mann–Whitney test. A nonparametric test was chosen because the *T* scores on the subscales did not follow a normal distribution. Data from ages 2 and 3 years were not analyzed because a different version of the CBCL with different subscales was used at those ages. Parent CBCL *T* scores at age 4 years were significantly higher among children with ADHD than among children without ADHD on all subscales (all $ps < .05$) except Somatic Complaints and Anxious/Depressed. At age 6 years, parent CBCL *T* scores were significantly higher among children with ADHD on all subscales (all $ps < .05$) except Somatic Complaints. Teacher CBCL *T* scores at that age were significantly higher among children with ADHD on all subscales (all $ps < .05$).

On the parent CBCL, at ages 9–11 years, children in the ADHD group had significantly higher *T* scores than children without ADHD on all eight subscales, including Attention Problems, ADHD $Mdn = 65$ versus non-ADHD $Mdn = 50$, $U = 5{,}903$, $z = -10.76$; Aggressive Behavior, ADHD $Mdn = 58$ versus non-ADHD $Mdn = 50$, $U = 10{,}453$, $z = -8.59$; and Delinquent Behavior, ADHD $Mdn = 62$ versus non-ADHD *Mdn* = 50, *U* = 11,638, *z* = −7.36, all *ps* < .001. On the teacher CBCL at ages 9–11 years, *T* scores on all subscales were also significantly higher among children with ADHD than without ADHD (all *ps* < .005).

Cognitive and receptive language measures. Figure 4 shows the mean standard scores for the ADHD and non-ADHD groups on the MSGCI at ages 3 and 4 years. At both ages, scores were lower (i.e., less favorable) in the ADHD group (at age 3, ADHD *M* = 92.03 ± 11.83 vs. non-ADHD *M* = 103.64 ± 12.81; at age 4, ADHD *M* = 86.35 ± 12.62 vs. non-ADHD *M* = 102.04 ± 13.75). Mean scores on the WISC–III at age 6 years were also lower in the ADHD group (88.98 ± 14.82 vs. 101.58 ± 14.01). Using repeated measures ANOVA, there was a main effect of group, $F(1, 641) = 66.83$, $p < .001$; this difference remained significant after adjusting for the same sociodemographic variables as those used in prior analyses, $F(1, 636) = 54.39$, $p < .001$. A significant Group × Time interaction also was found, $F(2, 640) = 6.74$, $p = .001$: From ages 3 to 4 years, mean standard scores on the MSGCI in children with ADHD showed a greater decrease than in children without ADHD.

Figure 5 shows the children's mean standard scores on the PPVT–R at ages 3, 4, and 6 years. Scores again were lower in children in the ADHD group than in the non-ADHD group. The main effect of group remained significant after adjusting for the same sociodemographic variables as those used in prior analyses, $F(1, 644) = 22.21$, $p < .001$. There was no significant Group × Time interaction, $F(2, 643) = 1.01$, $p = .366$.

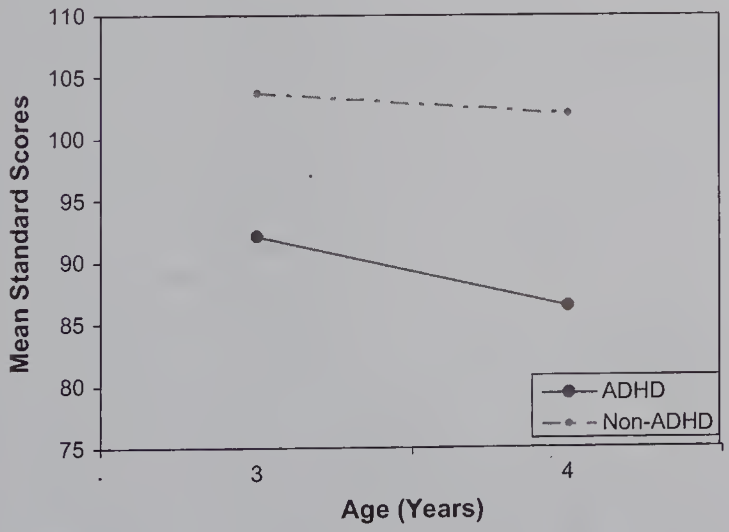


Figure 4. Mean standard scores on the General Cognitive Index of the McCarthy Scales of Children's Development in the children in the ADHD and non-ADHD groups at ages 3 and 4 years.
Note. ADHD = attention-deficit/hyperactivity disorder.

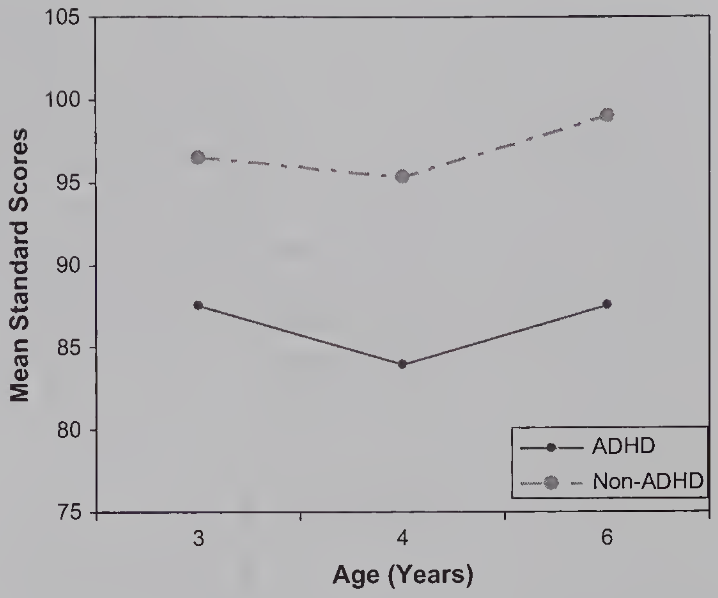


Figure 5. Mean standard scores on the Peabody Picture Vocabulary Test–Revised in the children in the ADHD and non-ADHD groups at ages 3, 4, and 6 years.
Note. ADHD = attention-deficit/hyperactivity disorder.

Impact of Learning Problem Status on Early Behavioral and Cognitive Measures in Children With and Without ADHD

Impact on early behavioral measures. Table 3 shows mean scores on the early childhood behavioral measures with the ADHD and non-ADHD groups divided into subgroups with and without learning problems. For the behavioral measures, a MANOVA was conducted using ADHD status and learning problems status as independent variables; type of health insurance, level of maternal education, and whether or not father was living in the home as covariates; and Total Problems *T* scores on the parent CBCL at 2, 3, 4, 6, and 9–11 years of age as the dependent variables. Results showed significantly higher (less favorable) Total Problems *T* scores in the ADHD group (with or without learning problems) at all ages, $F(4, 554) = 12.24$, $p < .001$, than in the learning problems group (with or without ADHD), $F(5, 554) = 1.45$, $p = .206$. There was no interaction between ADHD group and learning problems group, $F(5, 554) = 0.86$, $p = .511$.

Similar MANOVAs of teacher CBCL Total Problems *T* scores at ages 6 and 9–11 years showed a different pattern of results. Children with ADHD had higher scores than children without ADHD (regardless of learning problems status), $F(2, 627) = 18.54$, $p < .001$. Children with learning problems had higher scores than children than children without learning problems, $F(2, 627) = 5.81$, $p = .003$, and there was also a significant ADHD Group × Learning Problems Group interaction, $F(2, 627) = 5.75$, $p = .003$. These analyses included as covariates the same sociodemographic variables as those used in prior analyses.

Table 3

Behavioral and Cognitive Test Scores in Children Classified as ADHD With or Without Learning Problems

	ADHD											
	No						Yes					
	No learning problems (n = 531, 72%)			Learning problems (n = 141, 19%)			No learning problems (n = 22, 3%)			Learning problems (n = 47, 6%)		
Measures	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*
Parent CBCL (Total *T* scores)[a]												
Age 2 years	48.1	9.2	467	51.3	9.6	123	53.8	10.0	21	54.8	11.3	43
Age 3 years	47.9	9.5	480	50.1	8.8	115	56.2	6.6	20	54.0	11.7	43
Age 4 years	48.0	9.6	470	50.4	9.2	122	53.3	11.2	21	53.4	12.3	42
Age 6 years	45.6	9.6	477	49.4	10.6	122	55.6	8.1	20	55.9	11.4	43
Ages 9–11 years	46.0	10.0	410	51.7	11.2	99	61.0	9.3	19	60.8	11.9	39
Teacher CBCL (Total *T* scores)[a]												
Age 6 years	45.0	10.0	462	52.7	10.6	122	58.8	7.6	20	58.8	8.8	42
Ages 9–11 years	46.9	10.4	456	56.5	9.4	120	58.0	11.0	19	57.7	7.0	42
Cognition/intelligence/receptive language[b]												
MSGCI at age 3 years	105.1	12.8	490	95.9	11.3	121	97.8	11.5	20	89.3	11.1	42
MSGCI at age 4 years	104.0	13.1	490	92.7	12.9	125	94.9	8.3	21	82.4	12.5	44
WISC–III FSIQ	104.0	13.0	486	90.8	13.1	124	101.0	10.6	21	83.7	12.9	44
PPVT–R at age 3 years	98.0	15.4	496	89.1	12.3	122	93.2	8.4	20	84.6	15.0	41
PPVT–R at age 4 years	97.4	15.8	491	86.8	15.2	125	92.4	13.0	21	79.0	14.2	44
PPVT–R at age 6 years	101.0	14.7	484	88.9	13.4	125	94.9	11.4	21	83.5	14.5	44

Note. ADHD = attention-deficit/hyperactivity disorder; CBCL = Child Behavior Checklist; MSGCI = McCarthy Scales of Children's Abilities General Cognitive Index; WISC–III FSIQ = Wechsler Intelligence Scale for Children, Third Edition, Full-Scale Intelligence Quotient; PPVT–R = Peabody Picture Vocabulary Test–Revised, Form M.
[a] Higher scores reflect less favorable results. [b] Lower scores reflect less favorable results.

To probe these findings further using planned comparisons, an independent samples *t* test was conducted to compare Total Problems *T* scores on the parent CBCL at ages 2, 3, 4, and 6 years in the subgroups of children within the ADHD and non-ADHD groups who had and did not have learning problems. Using a Bonferroni correction because of multiple comparisons, significance was set at $p < .01$. In children with ADHD, the behavioral scores did not differ at age 2 years between the learning problems subgroup ($M = 54.77$, $SE = 1.73$) and the non-learning-problems subgroup ($M = 53.81$, $SE = 2.18$), $t(62) = -0.329$, $p = .743$, or at any of the later ages (all *p*s $> .05$). However, in children without ADHD, the behavioral scores did differ at age 2 years. Thus, non-ADHD children with learning problems had significantly higher Total Problems *T* scores ($M = 51.20$, $SE = 0.86$) than did non-ADHD children without learning problems ($M = 48.15$, $SE = 0.428$), $t(588) = 3.24$, $p = .001$. Non-ADHD children with learning problems also had significantly higher Total Problems *T* scores at age 6 than those without learning problems ($p < .001$).

Impact on early cognitive and receptive language measures. Similar analyses were conducted to assess the combined impact of ADHD status and learning problems status on cognitive and receptive language measures. For cognitive measures, a MANOVA was conducted using ADHD status and learning problems status as the independent variables; type of health insurance, level of maternal education, and whether or not the father was living in the home as covariates; and mean standard scores on the MSGCI at ages 3 and 4 years and on the WISC–III FSIQ scores at age 6 years as the dependent variables. At all ages, cognitive scores were significantly lower (less favorable) in the ADHD group compared to the non-ADHD group (regardless of learning problem status), $F(3, 632) = 8.00$, $p < .001$. Cognitive scores were also significantly lower at all ages in the learning problems group than in the non-learning-problems group (regardless of ADHD status), $F(3, 632) = 23.01$, $p < .001$. There was no interaction between ADHD group and learning problems group, $F(3, 632) = 1.34$, $p > .05$.

Regarding the receptive language measures, a MANOVA was conducted in which ADHD status and learning problems status were the independent variables; type of health insurance, level of maternal education, and whether or not the father was living in the home were covariates; and mean standard scores on the PPVT–M at ages 3, 4, and 6 years were the dependent variables. Mean scores at all ages were significantly lower in children with learning problems than in those without learning problems, $F(3, 640) = 14.34$, $p < .001$, but there were no significant differences in mean scores between children with ADHD and those without ADHD, $F(3, 640) = 1.43$, $p > .05$. There was no interaction between learning problems group status and ADHD group status, $F(3, 640) = 1.60$, $p > .05$. On both the cognitive and the receptive language measures, mean scores in children with both ADHD and learning problems were approximately 1 *SD* or more below the mean scores in children with neither ADHD nor learning problems.

Discussion

Main Findings

This study compared children 9–11 years of age with and without ADHD with regard to their behavioral, cognitive, and receptive language scores at 2, 3, 4, and 6 years of age and their behavioral scores at 9–11 years of age. Children in the ADHD group had consistently less favorable mean scores not only on behavioral ratings but also on cognitive and receptive language measures than did children in the non-ADHD group. Behavioral problem ratings in the ADHD group increased over time, whereas in the non-ADHD group, they decreased.

Prevalence of ADHD

The longitudinal study from which the present data are derived followed participants who were free at birth of many biological and social risk factors associated with ADHD. Despite those exclusion criteria, 9% of the sample had ADHD either reported or suspected and treated with medication. This value is within the range of 4%–12% reported in epidemiologic studies that do not exclude children with such risk factors (Brown et al., 2001) but higher than might be expected in a low-risk sample. Some of these children might not have met strict diagnostic criteria for ADHD, but the differences between groups on so many measures confirm that the two groups were substantially different from one another.

Prevalence of Learning Problems

Learning problems are highly prevalent in children diagnosed with ADHD. In our study, we classified children into the learning problems group on the basis of a parent's or a teacher's suspicion of a learning problem coupled with a child's receipt of any kind of remedial service. This liberal definition resulted in 68% of the children with ADHD having learning problems compared with 21% in the non-ADHD group. The reported rates of co-occurrence of learning disorders and ADHD vary across studies due to the lack of standardized definitions for learning disorders. The National Center for Health Statistics reported that in 1997 – 1998, 4% of U.S. children aged 6 – 11 years had both attention-deficit disorder and learning disability (Pastor & Reuben, 2002). Their definition of a specific learning disability comes from the Individuals With Disabilities Education Act, Section 300.541, and requires a discrepancy between intelligence and academic achievement (Individuals With Disabilities Education Improvement Act of 2004). This definition identifies fewer children than do other definitions. Other studies have reported the co-occurrence of ADHD and learning disorders to range from 11% (Wolraich, Hannah, Baumgaertel, & Feurer, 1998) to 70% (Mayes, Calhoun, & Crowell, 2000). Our results were similar to those of other studies that find substantial correlations between learning problems and ADHD (DuPaul & Stoner, 2003; Loe & Feldman, 2007) and, in particular, to the results reported by Biederman et al. (1998) who found that 84% of children with ADHD required extra help in school.

Characteristics of Children in the ADHD Group

We found that ADHD was associated with demographic and socioeconomic factors. As in other studies (Biederman et al., 2002; Gaub & Carlson, 1997), we found that larger proportions of children in the ADHD group were boys. Compared with children in the non-ADHD group, children with ADHD were also more likely to be enrolled in Medicaid insurance, to have parents with lower educational attainment, and to have fathers not living in the home. We considered type of health insurance and level of maternal education as indicators of SES, and whether the father was living in the home as an important demographic characteristic. The impact of SES on the diagnosis of childhood ADHD varies across studies. Although low family income and poverty status have been associated with lower probability of diagnosis (Bussing, Zima, Gary, & Garvan, 2003), a population-based cohort study found inverse associations of maternal and paternal education levels with risk for ADHD (St Sauver et al., 2004). A large, cross-sectional British epidemiologic survey found no independent association between ADHD diagnosis and either socioeconomic or family characteristics, including marital status and single-parent home (Ford, Goodman, & Meltzer, 2004). However, a study using a longitudinal cohort in the United States found that children with an ADHD diagnosis were less likely to be living with a biological father (Schneider & Eisenberg, 2006). The duration of MEE in this study differed between the ADHD and the non-ADHD groups; however, this difference did not persist after adjusting for demographic variables, presumably because the adverse demographic factors associated with ADHD are also important demographic risk factors for MEE.

Behavioral Measures

A major finding of this study was that the mean CBCL Total Problems *T* scores for children in the ADHD group were elevated in comparison to the children in the non-ADHD group as early as age 2 years. In addition, the Total Problems *T* scores increased (i.e., became less favorable) over the period between age 4 and 9 – 11 years, whereas the *T* scores for children in the non-ADHD group decreased during that period. Results were similar regarding the Internalizing and Externalizing *T* scores. Other studies of children developing typically report reductions over time in preschool behavioral problems such as aggression (Cummings, Ianotti, & Zahn-Waxler, 1989; Smith, Calkins, Keane, Anastopoulos, & Shelton, 2004). Our results show that children with ADHD were distinguishable from those without ADHD as toddlers and that the differences between the two groups become progressively more pronounced until at least school age.

Relations Between Behavioral Measures and Learning Problem Status

Another notable finding of this study was the increase in teacher ratings of behavioral problems in children in the learning problems group (without ADHD) from age 6 to 9 – 11 years. Parent reports showed that children in the learning problems group without ADHD had higher scores than children with neither ADHD nor learning problems. However, there was no Age × Group interaction. Teacher scores for the learning problems group at age 6 years were lower than those for children with ADHD with or without learning problems. However, the teacher scores became similar to those of children with ADHD with or without learning problems from age 6 to 9 – 11 years. This finding suggests that the number of behavioral problems in the children with learning problems and without a diagnosis of ADHD increases more in the school setting than in the home environment over time. Behavioral problems in this group of children may increase because of their struggles to meet the academic demands of the school.

Cognitive and Receptive Language Measures

Importantly, the findings of this study are consistent with the concept that ADHD is a cognitive as well as a behavioral disorder. The ADHD group had lower (less favorable) scores than the non-ADHD group on the MSGCI at ages 3 and 4 years and on the WISC – III FSIQ at age 6 years. These findings are consistent with those of other studies (Faraone et al., 1993; Marakovitz & Campbell, 1998). Children in the ADHD group also had lower mean scores on the PPVT – R, a test of receptive vocabulary, which correlates with verbal intelligence.

Similarly, children in the learning problems group had significantly lower cognitive and receptive vocabulary scores than children in the non-learning-problems group. The children with both ADHD and learning problems at school age had lower cognitive and receptive vocabulary scores at 3, 4, and 6 years of age than did children who had either one or none of the conditions. The differences were at least 1 *SD* in magnitude.

Limitations and Considerations for Future Studies

The estimates of ADHD prevalence in this study were based on parent reports of the diagnosis of ADHD or the use of relevant medication at ages 9–11 years. The prevalence of learning problems was based on a parent or teacher suspicion of learning problems and the receipt of remedial services of many different types. Future studies would ideally include more detailed diagnostic information using an ADHD-specific questionnaire or rating scale and a standardized diagnostic interview for ADHD and comorbid conditions. More detailed information from neuropsychological and/or psychoeducational testing could improve categorization of children's specific learning problems. Information on educational services (i.e., setting, provider, content, length, intensity) gathered by a structured services assessment measure such as the Services for Children and Adolescents – Parent Interview (Jensen et al., 2004) would also improve the ability to characterize remedial support for children's learning problems. Another limitation is the lack of information regarding family history of ADHD. Parent symptoms of ADHD are likely to affect parent – child interactions as well as how the parent views the child's behavior.

The exclusion criteria of the study from which the present sample was derived resulted in the exclusion of children at heightened risk for ADHD on the basis of low birth weight, prematurity, parental substance use, and mothers less than 18 years of age. The findings from our study may therefore not be generalizable to higher risk samples. In addition, the sample included relatively few Hispanic, Asian, and middle-class Black infants and no Native American infants (Paradise et al., 1997). Finally, parents who declined to enroll their infants tended to have higher SES than parents who enrolled their infants, and enrolled infants from lower SES families were more likely to have been lost to follow-up and to have incomplete data sets. On the other hand, upper- and lower-middle-class, working-class, and indigent families were all represented in substantial numbers (Paradise et al., 1997).

Developmental Implications and Conclusions

Research on the characteristics of children with ADHD during early childhood has been limited. This study provides support for the concept that ADHD is a chronic condition associated with adverse behavioral and cognitive characteristics in the toddler and preschool period. That the development of children with ADHD differs from children without ADHD at very young ages suggests that the disorder is not simply the result of poor adaptation to school environments. Behavioral and cognitive problems that are manifest in the preschool years in children who are ultimately diagnosed with ADHD in the middle school years appear to be relatively stable traits. Whereas in children who do not have ADHD, behavioral scores decrease (improve) over time, our data indicate that behavioral symptoms in children with ADHD tend to worsen by middle childhood, as rated by both parents and teachers.

In children with ADHD, the presence or absence of a learning problem did not appear to influence how parents or teachers viewed the level of behavioral problems. In contrast, in children with learning problems alone and not diagnosed with ADHD, teachers' perceptions of behavioral problems were much higher than parents' perceptions. This perception of increased behavioral problems at school age in the children with learning problems suggests that difficulty with academic work and poor adaptation to school are associated with deteriorating behavior over time.

Future longitudinal studies should attempt to determine whether early intervention for children at risk for ADHD can improve their long-term outcomes. In addition, the identification of other precursors of ADHD during the toddler and early preschool years may help enable the early identification of children with, or at risk for, ADHD and in turn might suggest interventions for improving their long-term academic, educational, and social outcomes.

References

Achenbach, T. (1991a). *Integrative guide for the 1991 CBCL/4-18, YSR, and TRF profiles*. Burlington: University of Vermont, Department of Psychiatry.

Achenbach, T. (1991b). *Manual for the Child Behavior Checklist/4-18 and 1991 profile*. Burlington: University of Vermont, Department of Psychiatry.

Achenbach, T. (1992). *Manual for the Child Behavior Checklist/2-3 and 1992 profile*. Burlington: University of Vermont, Department of Psychiatry.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

August, G. J., Braswell, L., & Thuras, P. (1998). Diagnostic stability of ADHD in a community sample of school-age children screened for disruptive behavior. *Journal of Abnormal Child Psychology, 26*, 345–356.

Barbaresi, W. J., Katusic, S. K., Colligan, R. C., Weaver, A. L., & Jacobsen, S. J. (2007). Long-term school outcomes for children with attention-deficit/hyperactivity disorder: A population-based perspective. *Journal of Developmental & Behavioral Pediatrics, 28*, 265–273.

Barkley, R. A. (2006). *Attention-deficit hyperactivity disorder: A handbook for diagnosis and treatment* (3rd ed.). New York: Guilford.

Bhutta, A. T., Cleves, M. A., Casey, P. H., Cradock, M. M., & Anand, K. J. S. (2002). Cognitive and behavioral outcomes of school-aged children who were born preterm: A meta-analysis. *Journal of the American Medical Association, 288*, 728–737.

Biederman, J., Faraone, S. V., Taylor, A., Sienna, M., Williamson, S., & Fine, C. (1998). Diagnostic continuity between child and adolescent ADHD: Findings from a longitudinal clinical sample. *Journal of the American Academy of Child & Adolescent Psychiatry, 37*, 305–313.

Biederman, J., Mick, E., Faraone, S. V., Braaten, E., Doyle, A., Spencer, T., et al. (2002). Influence of gender on attention deficit hyperactivity disorder in children referred to a psychiatric clinic. *American Journal of Psychiatry, 159*, 36–42.

Botting, N., Powls, A., Cooke, R. W. I., & Marlow, N. (1997). Attention deficit hyperactivity disorders and other psychiatric outcomes in very low birth weight children at 12 years. *Journal of Child Psychology and Psychiatry, 38*, 931–941.

Brown, R. T., Freeman, W. S., Perrin, J. M., Stein, M. T., Amler, R. W., Feldman, H. M., et al. (2001). Prevalence and assessment of attention-deficit/hyperactivity disorder in primary care settings. *Pediatrics, 107*, E43.

Bussing, R., Zima, B. T., Gary, F. A., & Garvan, C. W. (2003). Barriers to detection, help-seeking, and service use for children with ADHD symptoms. *Journal of Behavioral Health Services & Research, 30*, 176–189.

Camp, B. W. (1996). Adolescent mothers and their children: Changes in maternal characteristics and child developmental and behavioral outcome at school age. *Journal of Developmental & Behavioral Pediatrics, 17*, 162–169.

Campbell, S. B., Pierce, E. W., March, C. L., Ewing, L. J., & Szumowski, E. K.. (1994). Hard-to-manage preschool boys: Symptomatic behavior across contexts and time. *Child Development, 65*, 836–851.

Campbell, S. B., Pierce, E. W., Moore, G., & Marakovitz, S. (1996). Boys' externalizing problems at elementary school age: Pathways from early behavior problems, maternal control, and family stress. *Development and Psychopathology, 8*, 701–719.

Campbell, S. B., Shaw, D. S., & Gilliom, M. (2000). Early externalizing behavior problems: Toddlers and preschoolers at risk for later maladjustment. *Development and Psychopathology, 12*, 467–488.

Cummings, E. M., Ianotti, R. J., & Zahn-Waxler, C. (1989). Aggression between peers in early childhood: Individual continuity and developmental change. *Child Development, 60*, 887–895.

Dey, A. N., & Bloom, B. (2005). Summary health statistics for U.S. children: National Health Interview Survey, 2003. *Vital & Health Statistics—Series 10: Data From the National Health Survey, 223*, 1–78.

Dunn, L. (1981). *Peabody Picture Vocabulary Test–Revised: Manual for Forms L and M*. Circle Pines, MN: American Guidance Service.

DuPaul, G. J., McGoey, K. E., Eckert, T. L., & VanBrakle, J. (2001). Preschool children with attention-deficit/hyperactivity disorder: Impairments in behavioral, social, and school functioning. *Journal of the American Academy of Child & Adolescent Psychiatry, 40*, 508–515.

DuPaul, G. J., & Stoner, G. (2003). *ADHD in the schools: Assessment and intervention strategies* (Vol. 19, 2nd ed.). New York: Guilford.

Faraone, S. V., Biederman, J., Lehman, B. K., Spencer, T., Norman, D., Seidman, L. J., et al. (1993). Intellectual performance and school failure in children with attention deficit hyperactivity disorder and in their siblings. *Journal of Abnormal Psychology, 102*, 616–623.

Farrington, D. P., Loeber, R., & Van Kammen, W. B. (1990). Long-term criminal outcomes of hyperactivity-impulsivity-attention deficit and conduct problems in childhood. In L. N. Robins & M. Rutter (Eds.), *Straight and devious pathways from childhood to adulthood* (pp. 62–81). New York: Cambridge University Press.

Ford, T., Goodman, R., & Meltzer, H. (2004). The relative importance of child, family, school and neighbourhood correlates of childhood psychiatric disorder. *Social Psychiatry and Psychiatric Epidemiology, 39*, 487–496.

Gaub, M., & Carlson, C. L. (1997). Gender differences in ADHD: A meta-analysis and critical review. *Journal of the American Academy of Child & Adolescent Psychiatry, 36*, 1036–1045.

Gray, R., Indurkhya, A., & McCormick, M. (2004). Prevalence, stability, and predictors of clinically significant behavior problems in low birth weight children at 3, 5, and 8 years of age. *Pediatrics, 114*, 736–743.

Green, V. A., O'Reilly, M., Itchon, J., & Sigafoos, J. (2005). Persistence of early emerging aberrant behavior in children with developmental disabilities. *Research in Developmental Disabilities, 26*, 47–55.

Horwitz, S. M., Irwin, J. R., Briggs-Gowan, M. J., Heenan, J. M., Mendoza, J., & Carter, A. S. (2003). Language delay in a community cohort of young children. *Journal of the American Academy of Child & Adolescent Psychiatry, 42*, 932–940.

Individuals With Disabilities Education Improvement Act. (2004).

Ingram, S., Hechtman, L., & Morgenstern, G. (1999). Outcome issues in ADHD: Adolescent and adult long-term outcome. *Mental Retardation and Developmental Disabilities Research Reviews, 5*, 243–250.

Jensen, P. S., Eaton Hoagwood, K., Roper, M., Arnold, L. E., Odbert, C., Crowe, M., et al. (2004). The services for children and adolescents-parent interview: Development and performance characteristics. *Journal of the American Academy of Child & Adolescent Psychiatry, 43*, 1334–1344.

Lahey, B. B., Pelham, W. E., Loney, J., Lee, S. S., & Willcutt, E. (2005). Instability of the DSM–IV subtypes of ADHD from preschool through elementary school. *Archives of General Psychiatry, 62*, 896–902.

Lehn, H., Derks, E. M., Hudziak, J. J., Heutink, P., van Beijsterveldt, T. C., & Boomsma, D. I. (2007). Attention problems and attention-deficit/hyperactivity disorder in discordant and concordant monozygotic twins: Evidence of environmental mediators. *Journal of the American Academy of Child & Adolescent Psychiatry, 46*, 83–91.

Loe, I. M., & Feldman, H. M. (2007). Academic and educational outcomes of children with ADHD. *Journal of Pediatric Psychology, 32*, 643–654.

Loeber, R., & Dishion, T. (1983). Early predictors of male delinquency: A review. *Psychological Bulletin, 94*, 68–99.

Mannuzza, S., Klein, R. G., Bonagura, N., & Konig, P. H. (1988). Hyperactive boys almost grown up: II. Status of subjects without a mental disorder. *Archives of General Psychiatry, 45*, 13–18.

Mannuzza, S., Klein, R. G., & Moulton, J. L., III. (2003). Persistence of attention-deficit/hyperactivity disorder into adulthood: What have we learned from the prospective follow-up studies? *Journal of Attention Disorders, 7*, 93–100.

Marakovitz, S. E., & Campbell, S. B. (1998). Inattention, impulsivity, and hyperactivity from preschool to school age: Performance of hard-to-manage boys on laboratory measures. *Journal of Child Psychology and Psychiatry, 39*, 841–851.

Mayes, S. D., Calhoun, S. L., & Crowell, E. W. (2000). Learning disabilities and ADHD: Overlapping spectrum disorders. *Journal of Learning Disabilities, 33*, 417–424.

McCarthy, D. (1972). *McCarthy Scales of Children's Abilities*. San Antonio, TX: Psychological Corporation.

McCormick, M., Workman-Daniels, K., & Brooks-Gunn, J. (1996). The behavioral and emotional well-being of school-aged children with different birth weights. *Pediatrics, 97*, 18–25.

McGee, R., Partridge, F., Williams, S., & Silva, P. A. (1991). A twelve-year follow-up of preschool hyperactive children. *Journal of the American Academy of Child & Adolescent Psychiatry, 30*, 224–232.

Nolan, E. E., Gadow, K. D., & Sprafkin, J. (2001). Teacher reports of DSM–IV ADHD, ODD, and CD symptoms in schoolchildren. *Journal of the American Academy of Child & Adolescent Psychiatry, 40*, 241–249.

Paradise, J. L., Campbell, T. F., Dollaghan, C. A., Feldman, H. M., Bernard, B. S., Colborn, D. K., et al. (2005). Developmental outcomes after early or delayed insertion of tympanostomy tubes. *New England Journal of Medicine, 353*, 576–586.

Paradise, J. L., Dollaghan, C. A., Campbell, T. F., Feldman, H. M., Bernard, B. S., Colborn, D. K., et al. (2000). Language, speech sound production, and cognition in three-year-old children in relation to otitis media in their first three years of life. *Pediatrics, 105*, 1119–1130.

Paradise, J. L., Dollaghan, C. A., Campbell, T. F., Feldman, H. M., Bernard, B. S., Colborn, D. K., et al. (2003). Otitis media and tympanostomy tube insertion during the first three years of life: Developmental outcomes at the age of four years. *Pediatrics, 112*, 265–277.

Paradise, J. L., Feldman, H. M., Campbell, T. F., Dollaghan, C. A., Colborn, D. K., Bernard, B. S., et al. (2001). Effect of early or delayed insertion of tympanostomy tubes for persistent otitis media on developmental outcomes at the age of three years. *New England Journal of Medicine, 344*, 1179–1187.

Paradise, J. L., Feldman, H. M., Campbell, T. F., Dollaghan, C. A., Rockette, H. E., Pitcairn, D. L., et al. (2007). Tympanostomy tubes and developmental outcomes at 9 to 11 years of age. *New England Journal of Medicine, 356*, 248–261.

Paradise, J. L., Rockette, H. E., Colborn, D. K., Bernard, B. S., Smith, C. G., Kurs-Lasky, M., et al. (1997). Otitis media in 2253 Pittsburgh-area infants: Prevalence and risk factors during the first two years of life. *Pediatrics, 99*, 318–333.

Pardini, D., Obradovic, J., & Loeber, R. (2006). Interpersonal callousness, hyperactivity/impulsivity, inattention, and conduct problems as precursors to delinquency persistence in boys: A comparison of three grade-based cohorts. *Journal of Clinical Child & Adolescent Psychology, 35*, 46–59.

Pastor, P. N., & Reuben, C. A. (2002). Attention deficit disorder and learning disability: United States, 1997-98. *Vital & Health Statistics—Series 10: Data From the National Health Survey, 206*, 1–12.

Schneider, H., & Eisenberg, D. (2006). Who receives a diagnosis of attention-deficit/hyperactivity disorder in the United States elementary school population. *Pediatrics, 117*, e601–e609.

Smith, C. L., Calkins, S. D., Keane, S. P., Anastopoulos, A. D., & Shelton, T. L. (2004). Predicting stability and change in toddler behavior problems: Contributions of maternal behavior and child gender. *Developmental Psychology, 40*, 29–42.

Sonuga-Barke, E. J., Auerbach, J., Campbell, S. B., Daley, D., & Thompson, M. (2005). Varieties of preschool hyperactivity: Multiple pathways from risk to disorder. *Developmental Science, 8*, 141–150.

Stevenson, J., & Richman, N. (1978). Behavior, language, and development in three-year-old children. *Journal of Autism Developmental Disorders, 8*, 299–313.

St. Sauver, J. L., Barbaresi, W. J., Katusic, S. K., Colligan, R. C., Weaver, A. L., & Jacobsen, S. J. (2004). Early life risk factors for attention-deficit/hyperactivity disorder: A population-based cohort study. *Mayo Clinic Proceedings, 79*, 1124–1131.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children* (3rd ed.). San Antonio, TX: Psychological Corporation.

Weiss, G., & Hechtman, L. (1993). *Hyperactive children grown up: ADHD in children, adolescents, and adults* (2nd ed.). New York: Guilford.

Wolraich, M. L., Hannah, J. N., Baumgaertel, A., & Feurer, I. D. (1998). Examination of DSM–IV criteria for attention-deficit/hyperactivity disorder in a county-wide sample. *Journal of Developmental & Behavioral Pediatrics, 19*, 162–168.

Child Development, November/December 2008, Volume 79, Number 6, Pages 1869 – 1890

Postural and Object-Oriented Experiences Advance Early Reaching, Object Exploration, and Means – End Behavior

Michele A. Lobo and James C. Galloway
The University of Delaware

The effects of 3 weeks of social (control), postural, or object-oriented experiences on 9- to 21-week-old infants' (N = 42) reaching, exploration, and means – end behaviors were assessed. Coders recorded object contacts, mouthing, fingering, attention, and affect from video. Postural and object-oriented experiences advanced reaching, haptic exploration of objects, and developing means – end behavior compared to social experience. Object-oriented experience best-advanced means – end behavior. The results suggest that the development of novel behaviors is dependent on multiple subsystems and can be similarly advanced by addressing a variety of these subsystems. They also suggest that past experiences with active object exploration can facilitate early information processing and the development of early knowledge.

Early experiences shape the development of many skills in infancy, such as early stepping (Vereijken & Thelen, 1997; Zelazo, Zelazo, Cohen, & Zelazo, 1993), reaching (Lobo, Galloway, & Savelsbergh, 2004), object exploration (Needham, Barrett, & Peterman, 2002), and crawling (Adolph, Vereijken, & Denny, 1998). The goal of this project was to assess the effects of advanced parent – infant experiences on the future development of three behaviors proposed to be developmentally related: (a) reaching, (b) object exploration, and (c) means – end behavior. We predicted that altering parent – infant interactions when infants were 2 – 3 months of age and not yet reaching would propel infants onto different developmental trajectories in the months after the prescribed experiences ended. In the current study, parents provided just 3 weeks of advanced postural or object-oriented experiences. We assessed the effects of these experiences on future reaching, object exploration, and means – end behavior across a 12-week period.

The onset of the ability to reach at about 4 months of age is a milestone that alters daily life for infants and their parents (E. J. Gibson, 1988; Thelen et al., 1993; von Hofsten, 1991). Prior to the onset of reaching, infants in the United States spend most of their time supine on their backs engaging in face-to-face social interactions (Davis, Moon, Sachs, & Ottolini, 1998; Fogel, Messinger, Dickson, & Hsu, 1999; Mildred, Beard, Dallwitz, & Unwin, 1995). If parents introduce objects to their infants in these months, they typically present them to infants out of reach for visual inspection (Eppler, 1995; Fogel, 1997; Fogel et al., 1999). Parents occasionally place objects in infants' hands or mouths; however, the grasping ability required for sustained oral or manual object exploration does not emerge until after the onset of reaching (Konczak & Dichgans, 1997; Rochat, 1989). Infants move their arms often in the period before they begin to reach with and without the presence of objects or people. It is through these spontaneous movements that infants explore the biomechanics of their bodies and the forces acting upon them and gain the motor control required for later object interaction (Bhat, Heathcock, & Galloway, 2005; Thelen, 1990; von Hofsten, 1993). In effect, the period before the onset of reaching could be termed a primarily "supine, spontaneous, and social period."

After reaching emerges, infants' and parents' daily activities change. Infants move their arms with more control, which allows for more purposeful interaction with objects (Bhat et al., 2005; Thelen et al., 1993). The onset of reaching also coincides with the time when infants typically begin to assume more varied postural orientations, rolling prone and supine, being

---

This study was conducted as the doctoral thesis for M.A.L. and was funded by dissertation fellowships through The University of Delaware. We thank the families who generously agreed to participate in this study, our wonderful undergraduate research assistants, especially Rachel Farley, Rebecca Silver, Lindsey Moore, Morgan Randles, and Lauren Emerson, for hours of coding assistance and the members of the thesis committee, Nancy Getchell, Slobodon Jaric, and Mark Stanton, for their invaluable input throughout the project.

Correspondence concerning this article should be addressed to Michele A. Lobo, Motor Behavior Laboratory, Department of Physical Therapy, 329 McKinly Lab Building, The University of Delaware, Newark, DE 19716. Electronic mail may be sent to malobo@udel.edu.

© 2008, University of Delaware
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7906-0023

placed vertical and in semireclined sitting more often, and sitting with assistance (Fogel et al., 1999; Snow, 1989). Better postural control coupled with reaching and grasping ability allows infants to independently explore objects in new ways for sustained periods through mouthing and touching (Adolph, Eppler, Marin, & Wechsler-Clearfield, 2000; J. J. Gibson, 1979). Furthermore, the onset of reaching changes the parent–infant–object dynamic. As parents observe their infants' emerging ability to interact with objects, they begin to present repeated, daily opportunities for their infants to haptically interact with objects (Fogel, 1997; Fogel et al., 1999; Reed & Bril, 1996). Compared to the prereaching period, the period following the onset of reaching could be termed a primarily "postures, purpose, and play period."

In our previous work, we showed that advanced general movement experiences and object-oriented experiences advanced infants' abilities to reach for objects (Heathcock, Lobo, & Galloway, 2008; Lobo et al., 2004). Below we detail why we chose to expand on these results by assessing the effects of advanced postural and object-oriented experiences on not only reaching but also on later object exploration and means–end behavior.

The Role of Postural Experiences in the Ontogeny of Reaching

The first objective of this project was to test whether advancing the development of postural stability would advance the onset of reaching. Postural stability involves the ability to dynamically maintain the position of the body, or the center of mass, within one's stability limits (Shumway-Cook & Woollacott, 1995). The role of parents in promoting the development of postural stability in early infancy has been documented in several cross-cultural (Bril & Sabatier, 1986; Hopkins & Westra, 1988, 1989, 1990) as well as prospective (Davis et al., 1998; Hadders-Algra, Brogren, & Forssberg, 1996) studies. Young infants lack the mobility to independently transition between positions, thus the variety of positions and postural challenges they experience is largely a function of how parents handle them. Therefore, the emergence of postural stability is dependent not only upon the evolving abilities of infants but also upon parents' responses to these abilities.

Postural stability has been proposed as a requirement for reaching (Bertenthal & von Hofsten, 1998; Hopkins & Ronnqvist, 2002; Rochat & Goubet, 1995; Savelsbergh & van der Kamp, 1993). Developmentally, postural advancements accompanied improvements in reaching (Fallang, Saugstad, & Hadders-Algra, 2000; Fontaine & le Bonniec, 1988; Rochat, 1992; Spencer & Thelen, 2000). Reaching also improved in real time when external postural stability was provided (Amiel-Tison, 1985; Grenier, 1981; Rochat & Goubet, 1995). If postural stability facilitates reaching, then advancing the development of postural stability should advance the emergence of reaching (Bertenthal & von Hofsten, 1998; Hopkins & Ronnqvist, 2002; Rochat & Goubet, 1995). Being placed more often in postures, such as sitting and standing, has been shown to directly advance the emergence and control of these postures (Bril & Sabatier, 1986; Hopkins & Westra, 1988, 1989, 1990; Zelazo et al., 1993). This is the first study to attempt to assess the effects of advanced postural stability on another unpracticed behavior (reaching). The postural experiences prescribed in this project were non-object-oriented experiences aimed at advancing postural stability across a variety of positions. We hypothesized that infants who received enhanced postural experience would have advanced reaching emergence and ability compared to control infants.

The Role of Object-Oriented Experiences in the Ontogeny of Reaching

The second objective of this project was to test whether providing earlier than typical object experience would advance reaching. Parents play an important role in helping their infants learn to reach around 4–5 months of age by presenting them with objects to reach for and scaffolding their early reaching attempts (Fogel, 1997; Fogel et al., 1999; Reed & Bril, 1996). Young infants cannot independently seek out objects; thus, they rely on parents to present them with opportunities to interact with objects. Therefore, as with postural stability, the emergence of reaching is dependent on the evolving abilities of infants as well as parents' responses to these evolving abilities (Bakeman, Adamson, Konner, & Barr, 1990; Bruner, 1972; Landry & Chapieski, 1988).

Experience moving and interacting with objects has been proposed to be an important factor in the emergence of reaching (Bhat & Galloway, 2006; Fogel et al., 1999; Thelen et al., 1993). Specifically, reaching was advanced after 2 weeks of daily, object-oriented movement experiences provided 1–2 months prior to typical reach onset (Lobo et al., 2004). If experience haptically interacting with objects facilitates reaching, then advancing parent–infant interactions to incorporate objects earlier should again advance the emergence of reaching (Fogel, 1997; Fogel et al., 1999; Lobo et al., 2004; Reed & Bril, 1996). Here, we assessed the effects on reaching behavior across a longer time span. The object-oriented experiences prescribed in this project were aimed at advancing infants' reaching ability and haptic experience with objects in supine. We hypothesized that infants who received enhanced object experience would have advanced reaching emergence and ability compared to control infants.

The Role of Reaching Experience in the Ontogeny of Object Exploration

The third objective of this project was to test for a developmental connection between reaching and object exploration. That is, would infants in groups who reached early due to the effects of enhanced postural or object-oriented experiences also display advanced object exploration? Infants learn to extract information about the affordances and properties of objects through active object interaction (Campos et al., 2000; E. J. Gibson, 1988; J. J. Gibson, 1979). Not surprisingly, the emergences of reaching and object exploration are important milestones linked with abilities across multiple domains, including cognitive, social, and language (Bremner, 2000; E. J. Gibson, 1988, 2000; Needham, 2000; Piaget, 1952, 1954). The onset of reaching is rapidly followed by the ability to grasp objects for sustained periods (Wimmers, Savelsbergh, Beek, & Hopkins, 1998; Wimmers, Savelsbergh, van der Kamp, & Hartelman, 1998). This grasping ability coupled with increased opportunities to explore objects presented by parents may be important in transitioning infants into more frequent, purposeful, exploratory object interactions.

Although experience actively exploring objects has been viewed as necessary for infants to learn about the actions they can perform on objects (Bourgeois, Khawar, Neal, & Lockman, 2005; Case-Smith, Bigsby, & Clutter, 1998; Needham, Cantlon, & Holley, 2006), for infants less than 6 months of age, we know of only one study assessing reaching and exploration behaviors in combination (Needham et al., 2002). In that study, nonreaching infants given experience interacting with objects using sticky mittens with Velcro later showed greater object engagement and haptic exploration of objects. If experience actively exploring objects facilitates early knowledge about how to examine objects, then advancing the onset of reaching should advance object exploration behavior. This project aimed to determine the effect of advanced reach onset on the ontogeny of object exploration. We hypothesized that infants in groups expected to reach earlier (postural and object experience) would show more haptic exploration of objects earlier compared to control infants.

The Role of Object Experience in the Ontogeny of Means–End Behavior

The final objective of this project was to test for a connection between reaching, object exploration, and means–end behavior. That is, would infants who had been reaching and exploring objects longer also display advanced means–end behavior? By acting on objects, infants learn information about the match between their actions and external events (E. J. Gibson, 2000). For instance, by exploring a jack-in-the-box, infants learn that cranking the handle results in the sight and sounds of a toy. Such active interaction with objects has been proposed to be the vehicle by which infants gain knowledge about more complex action–environment relationships such as means–end relationships (Bremner, 2000; E. J. Gibson & Pick, 2000; Piaget, 1952, 1954; von Helmholtz, 1962). Means–end behavior involves acting on one object as the means to affect another object (Munakata, Bauer, Stackhouse, Landgraf, & Huddleston, 2002). Most means–end studies test infants older than 6 months of age with the general consensus that means–end behavior is present by 8 months of age (Funk, 2002; Lewis, Alessandri, & Sullivan, 1990; Munakata et al., 2002).

Several studies have investigated the emergence of means–end knowledge (Funk, 2002; Gowen, Goldman, Johnson-Martin, & Hussey, 1989; Menard, 2005). Visual attention paradigms have suggested that infants may possess means–end knowledge before they can demonstrate means–end behavior (Hofstadter & Reznick, 1996; Sommerville & Woodward, 2005b). It remains unclear how to reconcile the discrepancies between the behavioral and visual studies. One account for the discrepancies is the means–end deficit account, which claims that young infants may possess means–end knowledge but cannot demonstrate it because they cannot perform the tasks typically used to measure means–end behavior (Shinskey & Munakata, 2001).

To study the early emergence of means–end behavior, we designed a task in which the means were in line with infants' skills and the end was meaningful to young infants. The requirements of the means–end tasks used in other studies are too difficult for infants younger than 6 months of age (Hood, 2001; Menard, 2005). For example, infants younger than 6 months are not able to grasp and pull a towel or to grasp and lift an enclosure while reaching for and grasping an object behind the enclosure with the opposite hand (Bojczyk & Corbetta, 2004; Corbetta, Thelen, & Johnson, 2000; Fagard, 2000). Similar to studies with older infants, our means–end task required that infants coordinate their existing behaviors to perform in an unfamiliar task, acting on one object to affect another while demonstrating intention via interest in the "end" object (Munakata et al., 2002; Willatts, 1999). However, our task required motor, perceptual, and affect regulation abilities that young infants possess. If experience acting on objects facilitates the learning of means – end relationships, then advancing the onset of reaching and object exploration should advance emerging means – end behavior. We hypothesized that infants who had the longest duration of active object exploration, those who received object-oriented experiences prior to the onset of reaching, would show the most advanced means – end behavior.

Method

Participants

Healthy, full-term infants entered the study when they were 8 – 11 weeks of age and not yet reaching. Before Visit 1, 42 infants were randomly assigned to the social experience (control) group (age at Visit 1: M = 61 days, SD = 5 days, range = 54 – 77 days), the postural movement experience group (age at Visit 1: M = 62 days, SD = 5 days, range = 57 – 79 days), or the object-oriented movement experience group (age at Visit 1: M = 61 days, SD = 5 days, range = 55 – 74 days). Infants among groups were matched for gender. Each group contained 7 boys and 7 girls. We located families with infants through local birth announcements. Parents provided informed consent. All infants were from two-parent families. Two infants were African Americans, and the remaining infants were Caucasians.

Fourteen additional infants were excluded from the sample. Nine were excluded because of advanced reaching at Visit 1 evidenced by more than 10 object contacts during the reaching task. Excluding these infants was important because one objective of this study was to determine the effect of advanced experiences on the initial emergence and subsequent development of reaching. The remaining 5 infants were excluded because of parents' inability to provide home experience for more than 60% of the days during the 3-week prescribed home experience period. We requested that parents provide the experiences daily yet emphasized that missing occasional days would be an acceptable reality. Parents reported in journals how much experience was provided daily and were unaware of the 60% criterion. Please see the Results section for specifics on parental reports of provision of experiences. Participants received certificates of participation and DVD recordings of their infants' first and fifth visits as gifts upon completion of the study.

Procedure

The same experimenter visited families in their homes for six visits over a period of 12 weeks. Visits 1, 3, 4, 5, and 6 were separated by 3 weeks. Visits 1, 2, and 3 were separated by 1.5 weeks. We documented each session using two synchronized video recordings to provide right- and left-side views of infants for coding. A splitter was used to place both views on one screen, and a vertical interval time code generator was used to superimpose the hour, minute, second, and frame on the image.

Parents provided enhanced experiences daily only between Visits 1 and 3. Thus, parents were asked to provide the experiences for a total of 3 weeks when infants were between 2 and 3 months old. At each visit, we assessed infants' ratings on the Alberta Infant Motor Scale (AIMS) and performance in three tasks: (a) use of arm movements to reach, (b) use of arm movements to explore objects, and (c) use of arm movements to explore a means – end task.

Rating on the AIMS

At each visit, we observed infants' general motor development and rated them using the AIMS (Piper & Darrah, 1994; Piper, Pinnell, Darrah, Maguire, & Byrne, 1992). AIMS scores were used to ensure that the three groups were not different in their motor development at the initiation of the study and as an independent measure of whether the activities provided to the postural experience group actually resulted in advanced postural development. The AIMS is a valid and reliable developmental assessment tool that compares infants' motor performance to that of a normative sample to provide percentile rankings reflecting general motor development. It consists of observation of infants' weight bearing, posture, and antigravity movement in supine, prone, sitting with support, and standing with support.

Task 1: Using Arm Movements to Reach

At each visit, infants were provided opportunities to reach for an object while lying supine and sitting upright in a custom seat (Figure 1). In the seat, infants had their trunks secured but had freedom of movement of their arms. In each position, an object was held stationary in midline at the infant's chest level for six 30-s trials. The object distance was determined at the

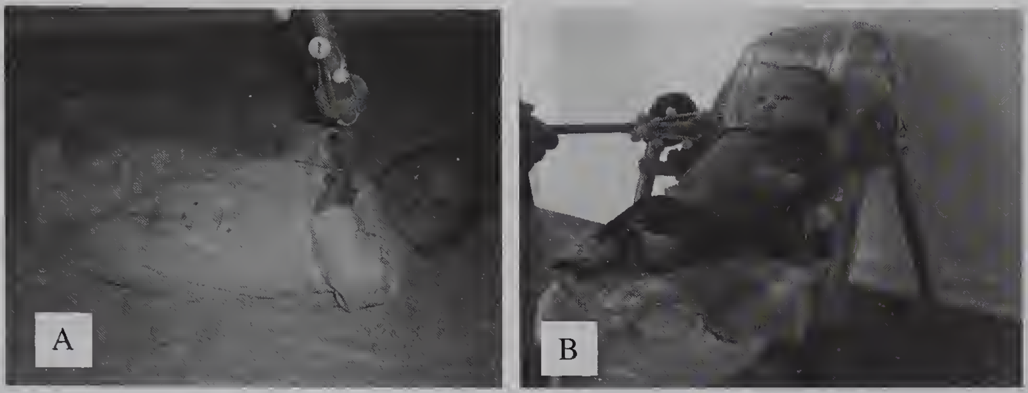


Figure 1. Task 1: One infant positioned supine (A) and vertically in the seat (B) for the assessment of the use of arm movements to reach.

beginning of each session by extending the infant's arm into midline, noting the location of the infant's wrist, and presenting the object in this location. Trials began when infants were in a positive behavioral state and visually attending to the object (Prechtl, 1974).

Task 2: Using Arm Movements to Explore Objects

At each visit, infants were provided opportunities to use arm and hand movements to haptically explore objects while sitting upright in our seat. Infants could not reliably grasp objects for prolonged periods at the start of the study. Thus, objects were attached to infants' wrists using Velcro bands (Figure 2). The experimental design for this task was based on Ruff (1984). Three pairs of objects varying in weight, texture, or sound-making ability were presented to infants for exploration. One object in each pair served as the familiarization object, and the other object served as the novel object. Infants had 60 s to explore each familiarization object and 45 s to explore each novel object. These times kept the familiarization period longer and novelty period shorter similar to Ruff. But the total presentation time for each pair (familiarization period plus novelty period) was shorter to increase the likelihood that the very young infants in this study could complete the entire paradigm. For the purposes of this article, we focus on global exploration of the objects by grouping the data.

At each visit, the familiarization and novelty trials were immediately preceded by the experimenter attaching the object to the wristband, raising the hand and object into the infant's view, and then releasing the hand at the start of the trial. There was no contact between the experimenter and infant during trials. Successive objects were attached to alternating hands. The initial hand of attachment was the right for half of the infants in each group and left for the others. Within infants, the initial hand of attachment was altered at each visit.

Task 3: Using Arm Movements to Explore a Means–End Task

At each visit, infants were provided opportunities to use arm movements to explore a means–end relationship while sitting upright in our seat (Figure 3). The means–end task began with a 30-s pre-means–end period followed by two means–end periods. During the pre-means–end period, a toy was placed to the left of the infant at a distance of 1 m (see C in Figure 3). The toy was within view but out of reach of the infant and was inactive. This period allowed us to

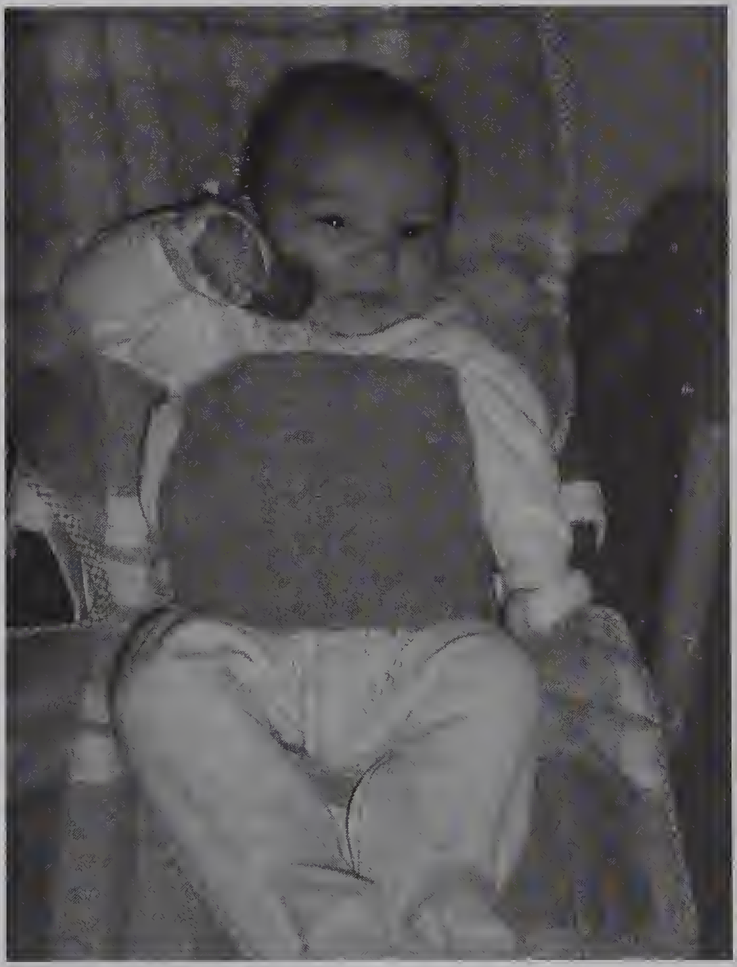


Figure 2. Task 2: One infant positioned in the seat for the assessment of the use of arm movements to explore objects.

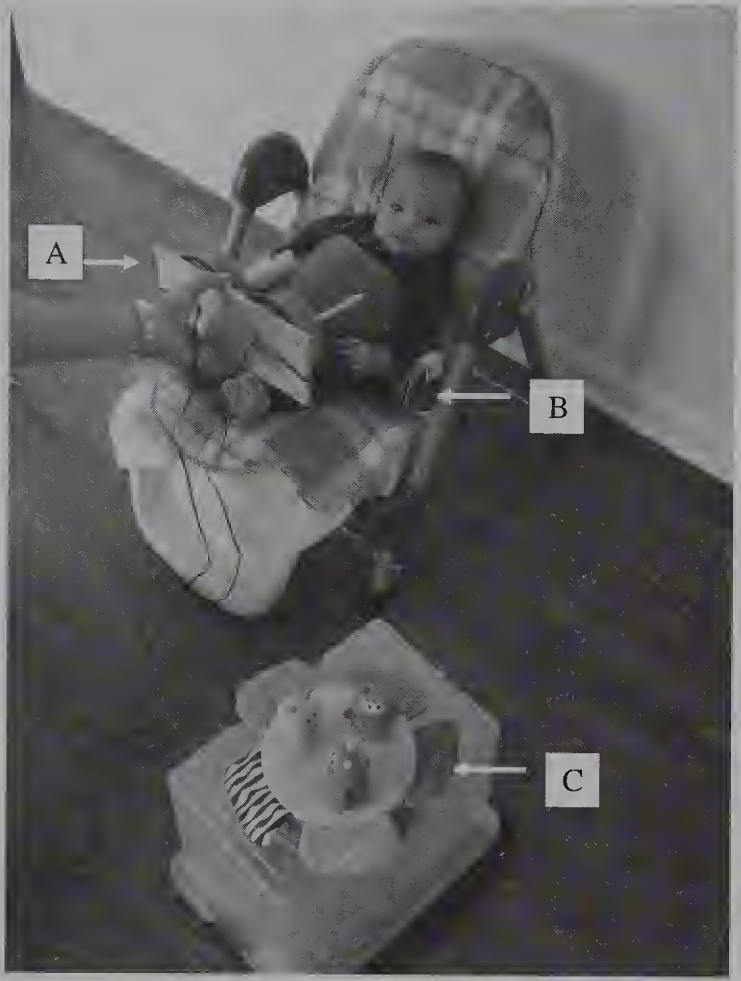


Figure 3. Task 3: One infant positioned in the seat for the assessment of the use of arm movements to explore a means – end task. *Note.* Note the two lever switches in the infant's reach (A), the small light secured to the chair (B; out of the infant's view), and the musical spinning bug toy (C; out of reach, in view for the infant).

test for changes in visual attention to the toy when it was not active or under infants' control.

The assessment continued with two means – end periods, one 4-min period followed by a second 3- to 4-min period. The length of the second period was dependent on the behavioral state of the infant (Prechtl, 1974). During both periods, the toy remained in its position and two lever switches were presented to infants within reach, one laterally to the right and one to the left (see A in Figure 3). To test infants' knowledge of the means – end relationship, switch – toy control was altered so one switch activated the toy during the first means – end period and the other switch activated the toy during the second means – end period. The position of the switch activating the toy during the first means – end period was counterbalanced among subjects in each group.

The lever switches were activated immediately when infants reached to displace them in any direction. One switch activated the toy ("toy activations"), which resulted in colorful plastic bugs spinning in circles to music. The other switch activated a light attached to the seat ("nonactivations"; see B in Figure 3). The light allowed coding of nonactivations and was out of the infant's view.

A certain level of activity was required for infants to demonstrate means – end behavior. Therefore, in the early visits, we discontinued the assessment if infants produced five or less total activations of the switches within the first 2 min of the first means – end period. When an infant met this criterion, the entire assessment was completed for that visit and all subsequent visits.

Prescribed Home Experiences

Parents in each group were asked to perform the home experiences for 15 min daily throughout the first 3 weeks of the study (Visit 1 to Visit 3) in addition to their typical daily activities. After Visit 3, parents were informed that they did not need to continue performing the home experiences for the purposes of this project. Although this amount of experience seems minimal, we have seen large changes in behavior as a result of prescribed experiences of similar durations (Lobo et al., 2004). It is likely that these changes result from carryover of the prescribed activities into daily parent – child interactions.

Parents were taught to provide the experiences via demonstrations from the experimenter at Visit 1 along with illustrated instruction manuals kept by parents. At Visit 2, parents were asked to perform an independent demonstration of the experiences with their infants for the experimenter and to discuss any questions or concerns regarding the experiences. Parents tracked the frequency and duration of their performance of the experiences in daily journals.

Social experience. Parents in the social experience group were asked to position their infants in supine while encouraging face-to-face, non-object-oriented interaction for 15 min. This served as a control for the increased social interaction and associated general movement that infants in the postural and object experience groups would receive.

Postural experience. Parents in the postural experience group were asked to perform non-object-oriented activities to improve postural stability and midline hand movement ability for 15 min. Specifically, the activities involved parents assisting infants to push up in prone, to pull up supine to sitting, to maintain their heads and upper trunks upright when their lower body was moved different directions in supported sitting and standing, and to move their hands to midline for clapping play.

Object-oriented experience. Parents in the object-oriented experience group were asked to teach their infants to reach for midline objects and to encourage haptic exploration of objects. To teach reaching, parents spent 5 min for each hand showing the infant his or her hand, showing an object at midline, moving the hand to the object, and then allowing the infant opportunities to perform this actively. To encourage haptic exploration of objects, parents spent 5 min assisting infants to feel objects varying in characteristics such as shape, size, texture, and hardness with their hands and mouths.

Data Analysis

Arm movement task coders were blind to group assignment. AIMS coding was performed by a pediatric physical therapist. Coding reliability was assessed based on a comparison of agreements and disagreements from each visit. For each variable, the number of behaviors coders agreed and disagreed on was determined and percentage of reliability was calculated using the equation: $[\text{agreed}/(\text{agreed} + \text{disagreed})] \times 100$. This is a strict measure of reliability (Angulo-Kinzler & Horn, 2001; Harris & Lahey, 1978; Lobo et al., 2004). Individuals were qualified for coding once they achieved and maintained inter- and intrarater reliabilities greater than 90%.

Variables for Rating on the AIMS

Changes in AIMS ratings across the span of the home experience (Visit 1 to Visit 3) and across the span of the study (Visit 1 to Visit 6) were analyzed in total and for the supine, prone, sitting, and standing subscales.

Variables for Task 1: Using Arm Movements to Reach

The variables for this assessment were analyzed in total with secondary analyses for the supine and vertical positions because previous research suggested that early vertical reaching abilities are more advanced than supine (Carvalho, Tudella, & Savelsbergh, 2007; Savelsbergh & van der Kamp, 1994).

Numbers of infants reaching. An infant was categorized as reaching if he or she contacted the object more than five times in 3 min of supine or vertical object interaction. These criteria are based on our past reaching studies (Lobo et al., 2004) with infants demonstrating that infants who have learned to reach perform this action repeatedly and reliably across all sessions after the onset of reaching with high numbers of object contacts.

Numbers and durations of object contacts. Object contact occurred when any surface of the hand contacted the object. Right- and left-limb variables were summed, such that, for example, one bilateral contact was counted as two contacts. Duration of each contact was coded using the time code on the video.

Numbers of infants reaching bilaterally and percent bilateral contact time. Bilateral midline reaching is typically observed after the onset of unilateral midline reaching (van Hof, van der Kamp, & Savelsbergh, 2002). We analyzed bilateral contact data because manipulating an object with two hands versus one allows infants to use actions not possible with just one hand to gather more information about objects' properties (Ballesteros, Manga, & Reales, 1997; Provine & Westerman, 1979). A bilateral contact was recorded any time right- and left-handed object contacts overlapped. An infant was recorded as reaching bilaterally if this occurred more than 10 times total or more than 5 times in supine or vertical. The total time of overlap between right- and left-handed contacts was extracted from the data and analyzed to determine the percentage of time infants were interacting with the object using both hands.

Variables for Task 2: Using Arm Movements to Explore Objects

The percent time fingering or mouthing the exploration objects was coded and calculated. "Fingering" occurred when the opposite hand contacted the attached object. "Mouthing" occurred when the mouth or lips contacted the attached object. The time spent fingering or mouthing was totaled and then converted to a percentage of the total possible exploration time or the sum of the trial times (three 60-s trials and three 45-s trials for a total of 315 s).

Variables for Task 3: Using Arm Movements to Explore a Means – End Task

Our means – end task was similar to the one used by Munakata et al. (2002) in that infants acted on a switch to inflict consequences on another, more distant object in the environment that did not appear physically linked to the switch. We modeled Willatts's (1999) use of multiple variables, including behavior with the means object and visual fixation on the end object, to assess for intentional means – end behavior. Means – end behavior in this paradigm was defined based on four criteria involving the three variables that follow.

1. *Number of toy activations and nonactivations.* Coders recorded toy and nonactivations noting the time superimposed on the video. These data provided (a) the total number of switch activations (toy activations plus nonactivations), (b) the number of toy activations, and (c) the number of nonactivations for each means – end period.
2. *Percent of toy activations and nonactivations accompanied by visual attention to the toy.* Coders determined whether infants visually attended to the toy within 5 s of each switch activation. Visual attention to the toy was not coded more than 5 s from an activation because each event of toy activity only lasted 5 s. These data were compiled to determine the percentage of toy activations and nonactivations accompanied by visual attention to the toy in each period.
3. *Percent time with positive or negative/neutral affect.* Each 5-s interval of each means – end period was coded using A System for Identifying Affect Expressions by Holistic Judgments (AFFEX), a general scale of affect for infants and children (Izard, Dougherty, & Hembree, 1989). Infants' affect on this scale could be coded as interest, happy, sad, angry, fearful, or neutral. No periods of fearful affect were observed, so affect was grouped into positive (interest or happy) or negative (sad or angry)/neutral. These data were compiled to determine the percent of time in each period during which infants demonstrated positive or negative/neutral affect.

Means – end behavior during one means – end period. The means – end task challenged young infants to coordinate their visual attention and arm movements to explore the relationship between switch activations and distant toy activity while maintaining a positive affect. An infant was recorded as having demonstrated means – end behavior in a period if he or she met four criteria incorporating the three aforementioned variables: (a) infants must first actively reach to explore, so infants were required to have an average of greater than two total activations per minute during the period; (b) infants should then match their actions to the desired toy activity, so it was required that greater than two thirds of the total activations be toy activations during the period (E. J. Gibson, 1997); (c) positive affect accompanies periods of learning, so it was required that infants demonstrate positive affect greater than two thirds of the total time during the period (Lewis et al., 1990; Lewis, Hitchcock, & Sullivan, 2004); and (d) as a suggestion of whether infants were able to link their actions with the toy activity, infants were required to look at the toy within 5 s of greater than one half of the toy activations (Willatts, 1999). To demonstrate that the criteria captured behavior that was unique to toy activations, the same criteria were also applied to non-activations.

Means – end behavior during both means – end periods. To allow for a period of exploration after the change in the switch – toy activation relationship between periods, only the final 2 min of the second means – end period were analyzed in cases where means – end behavior occurred in the first means – end period of the visit. We analyzed toy activations as well as nonactivations to ensure that the criteria captured behavior unique to activations resulting in toy activation. The four means – end criteria remained the same.

Secondary Variables Related to Means – End Behavior (Task 3)

Anticipated differences in means – end behavior among groups could potentially be attributed to differential development among groups. That is, groups may not have been able to achieve the four individual means – end behavior criteria in a similar manner due to differential motor, perceptual, or affect regulation abilities. We tracked the following secondary variables to determine whether any means – end behavior differences among groups were due to more general differences in the motor, perceptual, or affective requirements of the means – end task.

1. *Numbers of infants completing the entire assessment.* Differences in means – end behavior could have been due to differential motor ability to activate the switches. This may have resulted in group differences in the numbers of infants completing the entire means – end assessment. The numbers of infants completing the entire means – end performance assessment were recorded for each visit.
2. *Ability to turn to visually attend to the toy location.* Differences in means – end behavior could have been due to differential motor ability to rotate one's head to attend to the location of the toy. The ability to turn to visually attend to the toy included any instance of head rotation to attend to the toy during the means – end assessment or to the toy's anticipated location during the reaching and exploration assessments. The numbers of infants demonstrating the motor ability to turn to visually attend to the toy location were recorded for each visit.
3. *Visual attention to the toy or switches.* Differences in means – end behavior could have been due to differential visual – perceptual performance or

object interest during the means – end task. This may have resulted in group differences in visual attention to the switch and toy (means and end) and in the coordination of toy activations with attention to the toy. The data on visual attention to the toy or switches describe infants' looking behavior throughout the entire assessment, providing a broader picture of looking behavior than solely analyzing attention to the toy at times of switch activations. The durations infants looked at the toy during the 30-s pre-means – end period and at the toy or switches during the means – end periods at each visit were coded, summed, and converted to a percentage of the total period time.

Statistical Analyses

Kruskall – Wallace (KW) analysis of variance (ANOVA) was performed for most between-group comparisons. If the resultant α was $\leq .05$, KW post hoc tests by ranks were performed. Fisher's exact test was performed for planned between-group comparisons of nominal variables. For within-group comparisons of secondary means – end performance variables, Friedman's ANOVA was performed because the same participants contributed to the data being compared. In all statistical analyses, $\alpha \leq .05$ represents significant results. When multiple results are presented within the same set of parentheses, the first *p* value in a set represents the three-way result and the following *p* values represent the results of comparisons between pairs of groups.

Results

There were no differences among groups in terms of age, days between visits, or amount of home experience. Infants in each group received a similar number of days of home experience (social experience: $M = 18.0$, $SD = 2.7$; postural experience: $M = 19.1$, $SD = 1.7$; object experience: $M = 16.9$, $SD = 2.1$) and a similar number of total minutes of home experience (social experience: $M = 304.1$, $SD = 146.6$; postural experience: $M = 263.2$, $SD = 64.9$; object experience: $M = 252.5$, $SD = 41.2$). As mentioned, provision of home experience was sufficient for inclusion in the study if it occurred on more than 60% of the days between Visit 1 and Visit 3 according to journal reports from parents. We chose this criterion based on pilot data from our previous intervention studies in which infants who received experiences less than 50% of the time did not appear different from control infants (Lobo et al., 2004). This percentage was not significantly different among groups (social experience: $M = 87.5$, $SD = 11.5\%$; postural experience: $M = 88.1$, $SD = 5.9\%$; object experience: $M = 78.2$, $SD = 10.5\%$).

Rating on the AIMS

AIMS ratings were not significantly different among the three groups at Visit 1, suggesting that infants began the study with similar levels of general motor development. Changes in AIMS ratings suggested that the postural experiences effectively advanced motor development by advancing postural stability. This was particularly the case for prone skills after the 3-week home experience period compared to both other groups and at the end of the study compared to the social experience group. The change in total AIMS rating from Visit 1 to Visit 6 was different among the groups and was significantly greater for postural experience infants than for social experience infants, $\chi^2(2, N = 42) = 6.27$, $ps = .04, .02$, respectively. The change in prone subscale rating from Visit 1 to Visit 3 was different among groups and was significantly greater for the postural experience infants than for infants in the social experience and object experience groups, $\chi^2(2, N = 42) = 7.18$, $ps = .03, .02, .03$, respectively. The change in prone subscale rating from Visit 1 to Visit 6 continued to be different among groups with a greater change for postural experience infants compared to social experience infants, $\chi^2(2, N = 42) = 7.72$, $ps = .02, .01$, respectively.

Task 1: Using Arm Movements to Reach

Number of Infants Reaching

A greater number of postural and object experience infants reached earlier than social experience infants (Table 1). More object experience infants reached than did social experience infants at Visits 3 and 4 (*ps* = .02, .01, respectively). More postural experience infants reached than did social experience infants at Visit 3 ($p = .03$). When separating the data from the supine and vertical positions, these group differences were more evident in the vertical position earlier for both the postural and the object experience groups (Visit 2: *ps* = .08, .006, respectively; Visit 3: *ps* = .03, .01, respectively; Visit 4: $p = .02$ for object experience) than in the supine position (Visit 3: $p = .02$ for object experience; Visit 4: $p = .05$ for postural experience; Visit 5: $p = .002$ for object experience). The total number of infants reaching in

Table 1

Total Numbers of Infants Reaching Unilaterally and Bilaterally Across Visits for the SE, PE, and OE Groups

	Visit number					
Group	1	2	3	4	5	6
Numbers of infants reaching unilaterally						
SE	0	2	3	6	10	13
PE	0	4	9[a]	9	12	14
OE	0	5	10[a]	13[a]	13	14
Numbers of infants reaching bilaterally						
SE	0	0	0	0	5	12
PE	0	1	3	4[a]	11[a]	11
OE	0	0	2	4[a]	13[a]	13

Note. SE = social experience; PE = postural experience; OE = object experience.

[a] Signifies a difference from the SE group, $p \leq .05$.

each group was greater in the vertical position than the supine position for each group from Visits 2 through 6.

Number and Duration of Object Contacts

Postural and object experience infants contacted the object more frequently at Visits 3–5 compared to social experience infants (Figure 4a). There were differences among groups for average number of contacts at Visits 3, 4, and 5, Visit 3: $\chi^2(2, N = 42) = 8.30, p = .02$; Visit 4: $\chi^2(2, N = 42) = 8.32, p = .02$; Visit 5: $\chi^2(2, N = 42) = 8.39, p = .02$. Object experience infants contacted the object more frequently than did social experience infants at Visits 3, 4, and 5 (*ps* = .003, .004, and .01, respectively). Postural experience infants contacted the object more frequently than did social experience infants at Visit 4 ($p = .03$). When separating the data from the supine and vertical positions, these group differences were more evident in the supine position earlier, Visit 2: $\chi^2(2, N = 42) = 7.58, p = .02$; Visit 3: $\chi^2(2, N = 42) = 10.15, p = .01$; Visit 5: $\chi^2(2, N = 42) = 13.34, p = .001$, than in the vertical position, Visit 4: $\chi^2(2, N = 42) = 9.84, p = .01$. Yet, the average number of contacts in each group was greater in the vertical position for every group from Visits 2 through 6.

Postural and object experience infants contacted the object for longer durations at Visits 2–4 compared to social experience infants (Figure 4b). There were differences among groups for duration per contact at Visits 2, 3, 4, and 6, Visit 2: $\chi^2(2, N = 370) = 23.47, p = .000$; Visit 3: $\chi^2(2, N = 777) = 21.99, p = .000$; Visit 4: $\chi^2(2, N = 1169) = 20.58, p = .000$; Visit 6: $\chi^2(2, N = 3042) = 13.54, p = .001$. Object experience infants contacted the object longer than did social experience infants at Visits 2 and 4 (*ps* = .002, .000, respectively). Object experience infants contacted the object longer than did postural experience infants at Visits 4 and 6 (*ps* = .05, .001, respectively). Postural experience infants contacted the object longer than did social experience infants at Visits 2, 3, and 4 (*ps* = .000, .002, .001, respectively). Postural experience infants contacted the object longer than did object experience infants at Visits 2 and 3 (*ps* = .01, .000, respectively). Social experience infants contacted the object longer than did the postural experience infants at Visit 6 ($p = .002$). When separating the data from the supine and vertical positions, these group differences were more evident earlier in the vertical position, Visit 2: $\chi^2(2, N = 27) = 31.93, p = .000$; Visit 3: $\chi^2(2, N = 467) = 15.81, p = .000$; Visit 4: $\chi^2(2, N = 906) = 18.27, p = .000$, than in the supine position, Visit 3: $\chi^2(2, N = 467) = 15.81, p = .000$. The average duration of contact in each group was similar in both positions for every group for the early visits but was longer in supine for the final visit.

Number of Infants Reaching Bilaterally and Percent Bilateral Contact Time

A greater number of postural and object experience infants reached bilaterally earlier than social experience infants (Table 1). More postural and object experience infants had bilateral object contacts than did social experience infants at Visits 4 and 5 (Visit 4: *ps* = .05, .05, respectively; Visit 5: *ps* = .03, .002, respectively). When separating the data from the supine and vertical positions, these group differences were more evident in the vertical position earlier for both the postural and object experience groups (Visit 4: *ps* = .04, .006, respectively; Visit 5: $p = .02$ for object experience) than in the supine position (Visit 5: $p = .05$ for both groups). The total number of infants reaching bilaterally in each group was greater in the vertical position for every group from Visits 2 through 6.

Postural and object experience infants spent more time bilaterally contacting the object at Visits 4 and 5 (Figure 4c). There were differences among groups at Visits 4 and 5 in terms of the percent time infants were bilaterally contacting the object, Visit 4: $\chi^2(2, N = 42) = 10.24, p = .01$; Visit 5: $\chi^2(2, N = 42) = 7.64, p = .02$. Both postural and object experience infants spent more time bilaterally contacting the object than did social experience infants at these visits (Visit 4: *ps* = .04, .001, respectively; Visit 5: *ps* = .05, .002, respectively). When separating the data from the supine and vertical positions,

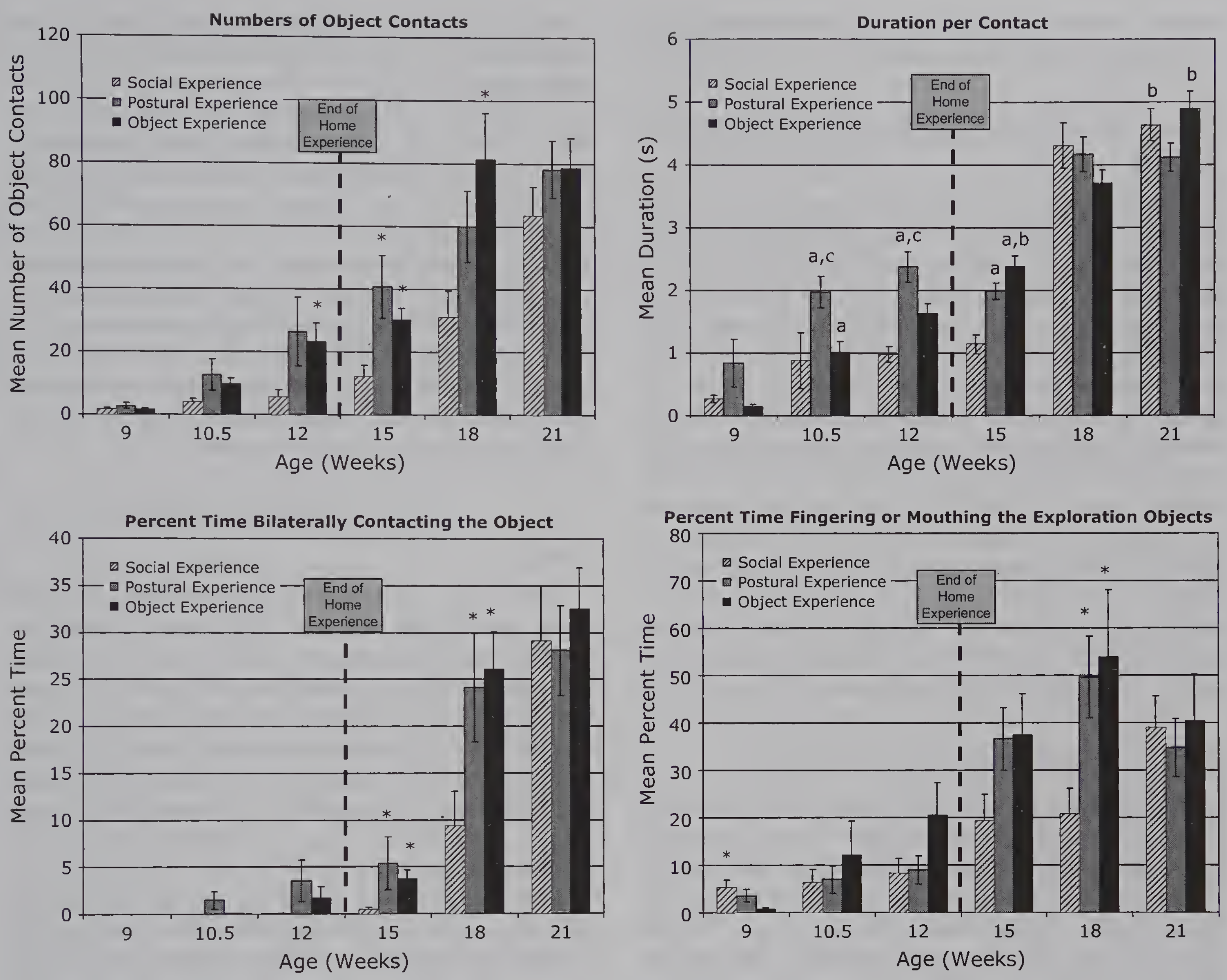


Figure 4. (a) Mean numbers of object contacts across visits for each group. (b) Mean durations per contact across visits for each group. (c) Mean percent times infants spent bilaterally contacting the object across visits for each group. (d) Mean percent times infants fingered or mouthed the exploration objects.

Note. Error bars reflect standard error of the mean.

[a]Signifies a difference from the social experience group. [b]Signifies a difference from the postural experience group. [c]Signifies a difference from the object experience group.

*Signifies a difference from the social experience group, $p \leq .05$.

these group differences were more evident in the vertical position earlier, Visit 4: $\chi^2(2, N = 42) = 9.60, p = .01$; Visit 5: $\chi^2(2, N = 42) = 6.08, p = .05$, than in the supine position, Visit 5: $\chi^2(2, N = 42) = 8.97, p = .01$. The average percent time spent bilaterally contacting the object in each group was greater in the vertical position for every group from Visits 2 through 6.

Task 2: Using Arm Movements to Explore Objects

Postural and object experience infants spent more time haptically exploring objects at Visit 5 than did social experience infants (Figure 4d). Infants in all groups increased the time they spent haptically exploring across the study, but at Visit 5, the postural and object experience groups spent more time fingering and mouthing the exploration objects than did the social experience group, $\chi^2(2, N = 42) = 8.20$, $ps = .02, .01, .04$, respectively. By the end of the study, infants in all groups were fingering and mouthing the exploration objects similarly.

Task 3: Using Arm Movements to Explore a Means – End Task

First, we report how each group performed during both means – end periods of a visit as well as the total number of periods in which infants displayed means – end behavior. We then report results from the secondary means – end variables.

Means – End Behavior During Both Means – End Periods of a Visit

Object experience infants demonstrated the best means – end behavior as evidenced by means – end behavior during both means – end periods of a visit. Table 2 shows the numbers of infants in each group demonstrating means – end behavior during both periods of the visit.

At Visit 6, more infants in the object experience group demonstrated means – end behavior during both periods of the visit than did infants in the social and postural experience groups (p = .049 for both comparisons). The object experience group also had the greatest total number of infants showing means – end behavior during both periods of a visit compared to the social experience group (p = .02). This means – end behavior was unique to toy activations. No infants in any group ever met the means – end behavior criteria for nonactivations during both periods of a visit.

Number of Periods With Means – End Behavior

Satisfying the means – end behavior criteria during just one means – end period of a visit does not demonstrate true means – end behavior. However, it does demonstrate emerging means – end behavior because this ability must be present in order for infants to show means – end behavior in both periods. Both the object and postural experience groups displayed more individual periods of means – end behavior than the social experience group. Specifically, the object and postural experience groups had greater numbers of total periods of means – end performance than the social experience group, $\chi^2(2, N = 42) = 6.31$, ps = .04, .02, .04, respectively. Object experience infants had 24 and postural experience infants had 21 periods of means – end behavior, whereas social experience infants had only 12 periods across all visits.

Table 2
Numbers of Infants Demonstrating Means – End Behavior During Both Means – End Periods of a Visit Across Visits and in Total for the SE, PE, and OE Groups

Group	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6	Total
SE	0	0	0	0	0	0	0
PE	0	0	0	0	1	0	1
OE	0	0	0	0	1	4[a,b]	5[a]

Note. SE = social experience; PE = postural experience; OE = object experience.
[a]Signifies a difference from the SE group, $p \leq .05$. [b]Signifies a difference from the PE group, $p \leq .05$.

The total numbers of periods where infants met these criteria for nonactivations were significantly lower for all groups (social experience, 1 period; postural experience, 2 periods; object experience, 0 periods) than were the number of periods where infants met these criteria for toy activations, social experience: $\chi^2(1, N = 28) = 11.63$, $p = .001$; postural experience: $\chi^2(1, N = 28) = 16.82$, $p = .000$; object experience: $\chi^2(1, N = 28) = 20.96$, $p = .000$, and were not different among groups. This suggests that the performance criteria captured behaviors unique to exploration of the switch – toy means – end relationship.

Results for Secondary Variables Related to Means – End Behavior

This means – end task required the coordination of motor, perceptual, and affective aspects of behavior. Results from secondary measures help to rule out whether the means – end behavior results simply reflect group differences in these requirements. Ruling out these general differences would provide further support that means – end behavior differences resulted from varying abilities of the infants in each group to coordinate these abilities to perform a challenging task.

Motor requirements for the means–end task. Group differences in means – end behavior were not explained by differences in motor ability. Infants in each group developed the motor requirements for the means – end performance task similarly. Motor requirements for the task involved the ability to (a) activate the switches using arm movement and (b) turn to the side to view the toy using head movement.

1. *Switch activation*: Each group activated the switches enough to have similar opportunities to display means – end behavior. Specifically, there were no significant differences among groups in the number of infants meeting the switch activation criterion more than five switch activations in the first 2 min of the first means – end period) and subsequently completing the entire assessment at each visit. At least two thirds of infants in each group completed the entire assessment by Visit 3, and all infants did so by Visit 6.

In addition, each group activated the switches a similar number of times at each visit. Specifically, there were no significant differences among groups for the number of toy activations or nonactivations except at Visit 4. Postural and object experience infants had more toy activations and nonactivations at Visit 4 than did social experience infants, toy activations: $\chi^2(1, N = 28) = 7.45$, $ps = .02, .01, .05$, respectively; nonactivations: $\chi^2(1, N = 28) = 8.86$, $ps = .01, .01, .02$, respectively.

2. *Head turning to visually attend to the toy*: Each group had the ability to turn their head to the side where the toy was located at each visit. There were no significant differences among groups in the numbers of infants who turned their heads toward the toy location. The majority of infants in each group demonstrated this ability at Visit 1, and all infants in each group demonstrated this ability by Visit 3.

Perceptual requirements for the means–end task. Group differences in means–end behavior were not due to general differences in visual–perceptual ability, although there were some differences among groups. That is, infants in each group developed the perceptual requirements for the means–end task similarly. These requirements were the ability to visually attend to (a) the toy during the pre-means–end period and (b) the toy or switches during the means–end periods.

Each group increased its visual attention to the toy during the pre-means–end period across visits. Each group had similar ability and interest to maintain attention to the toy during the pre-means–end period at each visit except at Visit 6. At Visit 6, the groups were different with object experience infants attending more to the toy than infants in the postural experience group, $\chi^2(2, N = 42) = 5.79$, $ps = .05, .02$, respectively.

Each group increased its visual attention to the toy and switches during the means–end periods across visits. Each group had similar ability and interest to maintain attention to the toy or switches during the means–end periods at each visit except at Visit 6. At Visit 6, the groups were different and the object experience group spent more time looking at the toy than did the social and postural experience groups, $\chi^2(2, N = 42) = 7.70$, $ps = .02, .01, .04$, respectively.

It was expected that once infants understood the switch–toy relationship, they would spend increased time looking at the toy, a focal point of the means–end behavior (Willatts, 1999). Therefore, we wanted to analyze not only how much each group attended to the toy compared to the other groups but also how much time within each group involved looking at the toy versus the switches at each visit. The social and postural experience groups spent more time looking at the switches than the toy at every visit (Figure 5a), Visit 2: social experience $\chi^2(1, N = 14) = 10.00$, $p = .002$, postural experience $\chi^2(1, N = 14) = 9.31$, $p = .002$; Visit 3: social experience $\chi^2(1, N = 14) = 6.23$, $p = .01$, postural experience $\chi^2(1, N = 14) = 9.31$, $p = .002$; Visit 4: social experience $\chi^2(1, N = 14) = 13.00$, $p = .000$, postural experience $\chi^2(1, N = 14) = 14.00$, $p = .000$; Visit 5: social experience and postural experience $\chi^2(1, N = 14) = 14.00$, $p = .000$; Visit 6: social experience $\chi^2(1, N = 14) = 10.29$, $p = .001$, postural experience $\chi^2(1, N = 14) = 14.00$, $p = .000$. Infants in the object experience group showed this same pattern of looking until Visit 6, the visit where they demonstrated the greatest means–end performance, when they visually attended to the switches and the toy the same amount of time, $\chi^2(1, N = 14) = 10.29$, $p = .001$ for Visits 1–5.

Affect requirement for the means–end task. Group differences in means–end behavior were not explained by general differences in ability to regulate affect across visits. Each group met the means–end behavior affect requirement of maintaining a positive affect more than two thirds of the time in a period for the final three to four visits. Each group increased the time it maintained a positive affect across visits, and the duration of positive affect was not different among groups at each visit. Each group demonstrated more positive than negative/neutral affect from Visits 4 to 6 (Figure 5b), Visit 4: social experience, postural experience, and object experience $\chi^2(1, N = 14) = 7.14$, $p = .008$; Visit 5: social experience $\chi^2(1, N = 14) = 4.57$, $p = .03$, postural experience and object experience $\chi^2(1, N = 14) = 10.29$, $p = .001$; Visit 6: social experience and postural experience $\chi^2(1, N = 14) = 10.29$, $p = .001$, object experience $\chi^2(1, N = 14) = 14.00$, $p = .000$.

Motor–perceptual coupling requirement for the means–end task. Group differences in means–end behavior were explained, in part, by differences in motor–perceptual coupling abilities. Specifically, groups differed in their ability to meet the motor–perceptual coupling requirement of accompanying toy activations with visual attention to the toy.

The postural and object experience groups demonstrated selective motor–perceptual coupling behavior earlier than the social experience group. Groups did not differ in the percentage of toy activations accompanied by attention to the toy at each visit. However, true means–end behavior requires that infants do not simply randomly attend to the toy with any switch activation but rather that they selectively attend to the toy more often with toy activations than with nonactivations. We term this *selective*

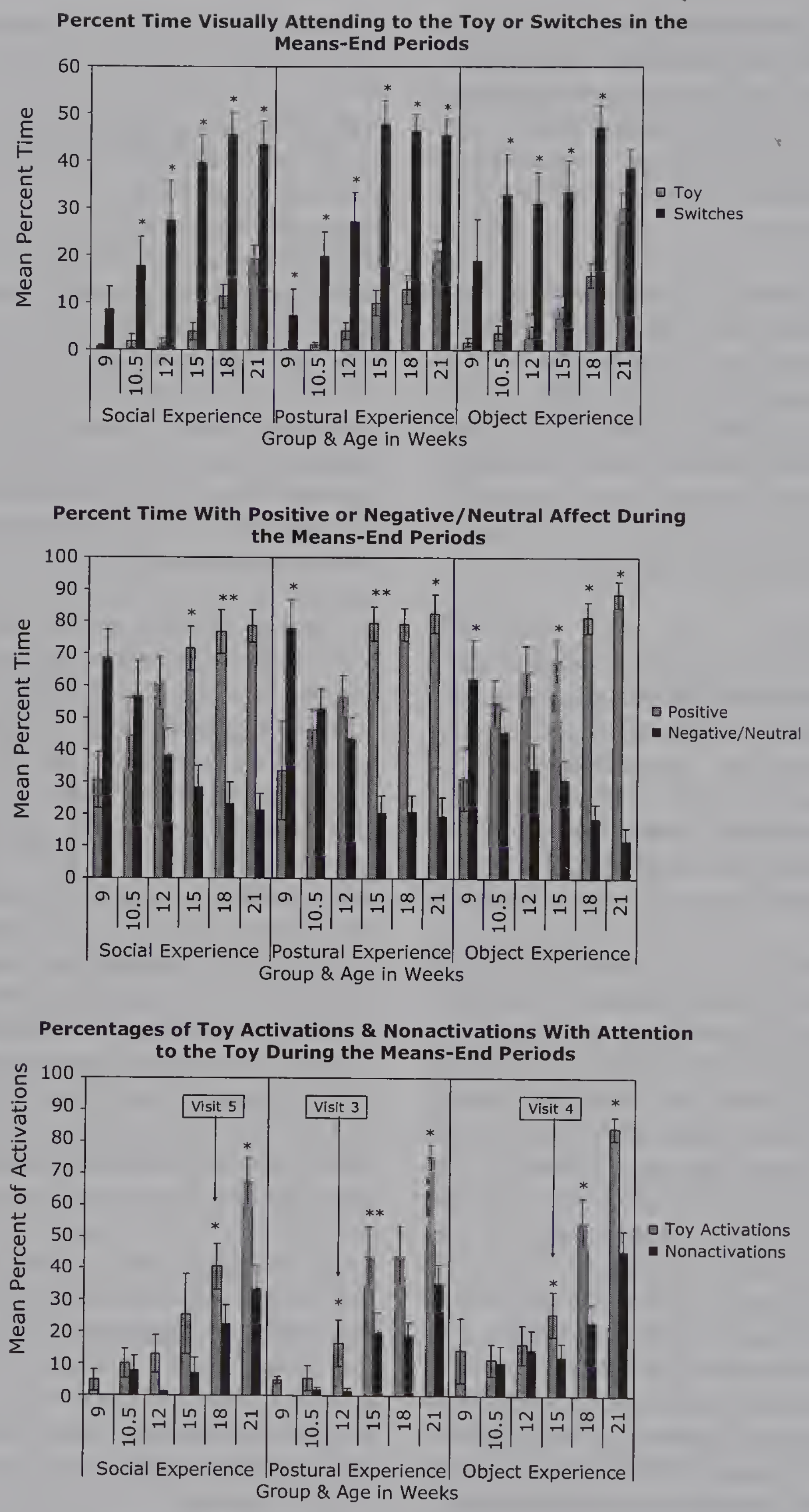


Figure 5. (a) Mean percent times each group attended to the toy or switches during the means–end periods across visits. (b) Mean percent times each group demonstrated positive or negative/neutral affect during the means–end periods across visits. (c) Mean percentages of toy activations and nonactivations for each group accompanied by attention to the toy during the means–end periods across visits.
Note. In Figure 5(a), note that the groups looked more at the switches except at Visit 6 when the object experience group looked similar amounts at the toy and switches. In Figure 5(b), note that each group increased positive affect and decreased negative/neutral affect over time without differences among groups. In Figure 5(c), note that the postural and object experience groups showed earlier selectivity of their motor–perceptual coupling than did the social experience group, attending to the toy more with toy activations than with nonactivations. Error bars reflect standard error of the mean.
$^*p \leq .05$.

motor–perceptual coupling. Indeed, group differences in this selective coupling emerged with greater attention to the toy with toy activations for postural experience infants by Visit 3 (Figure 5c), Visit 3: $\chi^2(1, N = 14) = 7.00, p = .01$; Visits 4 and 5: $\chi^2(1, N = 14) = 8.33, p = .004$; Visit 6: $\chi^2(1, N = 14) = 14.00, p = .000$, for object experience infants by Visit 4, Visit 4: $\chi^2(1, N = 14) = 8.33, p = .004$; Visit 5: $\chi^2(1, N = 14) = 9.308, p = .002$; Visit 6: $\chi^2(1, N = 14) = 14.00, p = .00$, and for social experience infants by Visit 5, Visit 5: $\chi^2(1, N = 14) = 10.00, p = .002$; Visit 6: $\chi^2(1, N = 14) = 7.14, p = .008$. In effect, the postural and object experience groups attended more often to the toy selectively with toy activations earlier than did the social experience group. This suggests that they had an earlier association between their actions on the switches and the activity of the toy.

Discussion

Summary of the Results

Postural and object experience infants had advanced reaching, exploration, and developing means–end behavior compared to social experience infants (Figure 6). Most of the group differences were observed after the 3-week prescribed home experience period ended at Visit 3. Differential ability to reach for and contact objects was the only change observed during the 3-week home experience period. Differences in reaching ability continued, and differences in object exploration and means–end behavior emerged weeks to months after the home experience period ended. Therefore, infants who had similar abilities at the first visit embarked on different developmental trajectories even after their prescribed home experience period ended. Specifically, postural and object experience infants had more frequent object contacts at Visits 2–5, longer contact durations at Visits 2–4, more bilateral contact time at Visits 4 and 5, more haptic object exploration at Visit 5, and greater numbers of periods of developing means–end behavior primarily at Visits 5 and 6 compared to social experience infants. The object experience group had the greatest number of infants demonstrating means–end behavior in both periods at Visit 6.

Task 1: Postural and Object Experience Infants Had Advanced Reaching Emergence and Ability

Postural and object experience infants had earlier onsets for reaching and advanced reaching ability compared to social experience infants. The differences in reaching ability among the three groups suggest that the emergence of reaching is dependent on multiple interacting subsystems and can be similarly advanced by experiences that affect a variety of these subsystems (Bertenthal & von Hofsten, 1998; Hopkins & Ronnqvist, 2002; Thelen, 1990; Thelen et al., 1993; von Hofsten, 1993). This supports and extends our previous work showing that the emergence of reaching can also be advanced via experiences to increase arm flapping (Lobo et al., 2004).

This study is the first to demonstrate that reaching emergence can be advanced by the provision of advanced postural experiences. This builds on previous work showing that experience being placed in different postures, such as sitting or standing, can

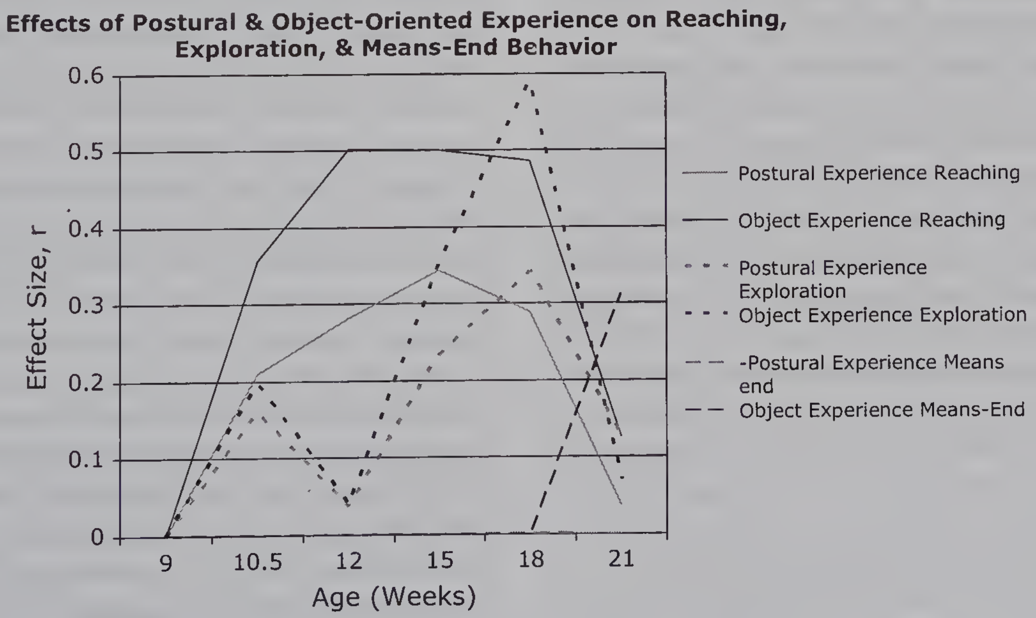


Figure 6. Effects of postural and object-oriented experiences on reaching, exploration, and means–end behavior represented by effect size, *r*.
Note. A value of zero indicates no difference in behavior from the social experience group.

directly advance infants' abilities to maintain these postures (Bril & Sabatier, 1986; Hopkins & Westra, 1988, 1989, 1990; Zelazo et al., 1993). This work was unique, however, in that the goal of the experiences was to directly advance postural stability, whereas our focus was whether this would then advance the unpracticed behavior of reaching.

To understand how postural experiences led to the emergence of reaching, we must consider factors both internal and external to the infants. Parents of postural experience infants were asked to provide infants with opportunities to improve their postural stability by placing and moving them in a variety of postures. This occurred during the supine, spontaneous, and social period when infants spend much of their time supine (Davis et al., 1998; Mildred et al., 1995). Ratings on the AIMS suggest that postural experience advanced infants' postural stability across postures. Postural experience infants likely had improved head and trunk strength and control for visual exploration and reaching, more appropriate muscle responses to maintain balance, and increased perceptual – motor experiences in a variety of positions (Bril & Sabatier, 1986; Thelen & Spencer, 1998; Wijnroks & van Veldhoven, 2003; Witherington et al., 2002). In addition, we suspect that as parents observed their infants' abilities to maintain different postures and to begin to reach for objects, they likely switched their parent – infant – object interactions to "postures, purpose, and play" earlier than typically reported (Fogel, 1997; Fogel et al., 1999; Reed & Bril, 1996). This switch was likely influenced by the effects of the postural experiences but was not assigned as part of the prescribed experiences. As a result, postural experience infants likely had early, increased opportunities to improve their postural stability across positions and to interact with objects throughout each day (Bertenthal & von Hofsten, 1998; Eppler, 1995). It is likely that such changes within infants combined with changes in parent handling and interaction to advance reaching ability in infants provided with postural experiences.

To understand how object-oriented experiences led to the emergence of reaching, we must again consider factors both internal and external to the infants. Parents of object experience infants were asked to provide their infants with opportunities and assistance to reach for and explore objects (Fogel, 1997; Fogel et al., 1999; Lobo et al., 2004). This occurred during the supine, spontaneous, and social period when parents primarily interact with infants face-to-face with little active object exploration (Fogel et al., 1999). Object experience infants likely had improved movement control in a midline task space, motivation to reach for objects presented, and knowledge that objects presented within reach afford haptic exploration. In addition, as parents observed their infants' abilities to reach for, grasp, and mouth objects, they likely switched their parent – infant – object interactions to "purpose and play" earlier than typical (Fogel, 1997; Fogel et al., 1999; Reed & Bril, 1996). Consequently, object experience infants likely had increased opportunities to interact with objects on a daily basis even outside of the prescribed home experience periods (Bakeman et al., 1990; Bruner, 1972). The home experience for these infants was provided in supine, and these infants did not advance in their postural stability according to AIMS ratings as did postural experience infants. In fact, their changes in total and subscale AIMS scores were not different from social experience infants. This suggests that parents of object experience infants did not provide infants the same advanced opportunities to improve their postural stability across positions as did parents of postural experience infants.

Social experience infants did not have advanced reaching ability. By Visit 6, however, their reaching abilities were similar to infants in the two advanced experience groups. The experiences parents of these infants provided involved face-to-face interactions in supine, the type of interaction most common during the supine, spontaneous, and social period (Davis et al., 1998; Fogel et al., 1999; Mildred et al., 1995). Therefore, it was expected that reaching development in these infants would mirror typical development. In fact, social experience infants began reaching at an average of 18.25 ± 1.81 weeks of age, which is within the 4- to 5-month-old period typically reported (Thelen et al., 1993; von Hofsten, 1991). Consequently, social experience infants likely did not transition to postures, purpose, and play and have increased opportunities to interact with objects on a daily basis until after their onset of reaching at Visit 5 (Eppler, 1995; Fogel, 1997; von Hofsten, 1997).

More infants were able to contact the object more frequently with one hand as well as bilaterally earlier and throughout the study in the vertical position compared to supine. The average duration per contact was not different between the vertical and supine positions throughout the study but was greater in the supine position by the end of the study. These data support the proposal that it is easier for young infants to initiate reaches but more difficult to sustain contact with objects in a vertical position (Carvalho et al., 2007; Savelsbergh & van der Kamp, 1994). In supine, infants may find it more challenging to initiate reaches but easier to sustain contacts with objects due to the effects of gravity on the arm's biomechanics. When the hand is farther away from the center of mass, infants must produce more muscle force to overcome gravitational force. This is the case when reaches are initiated in supine with the hands resting laterally on the floor and in trying to sustain contacts in vertical with the hands held in front of the trunk.

In this study, two distinct experiences advanced reaching ability. These results support the idea that even the earliest behaviors emerge as the result of a dynamic process involving the coordination of multiple interacting subsystems encompassing task, environmental, and organismic factors (Bertenthal & von Hofsten, 1998; Corbetta & Bojczyk, 2002; Hopkins & Ronnqvist, 2002; Joh & Adolph, 2006; Lee, Liu, & Newell, 2006; Thelen, 1990; Thelen et al., 1993; von Hofsten, 1993). Acceleration of the development of one or more of these subsystems can advance the ontogeny of early object interaction behavior (Lobo et al., 2004). For this study, we suspect that advancing parent – infant – object interactions, head and trunk control, postural stability, and / or perceptual – motor experiences across positions may have advanced reaching development.

Task 2: Postural and Object Experience Infants Had Advanced Haptic Exploration of Objects

Postural and object experience infants demonstrated advanced haptic exploration of objects compared to social experience infants. These results suggest that haptic exploration of objects is dependent on infants' past experiences, the emergence of reaching, as well as ongoing daily parent – infant – object interactions (Berthier, Rosenstein, & Barto, 2005; Bourgeois et al., 2005; J. J. Gibson, 1962, 1966; Millar, 2005; Needham, 2000; Needham et al., 2002). This is the first study presenting empirical evidence for a link between the onset of reaching and later haptic object exploration across developmental time. It is important to note that increased haptic exploration of objects did not occur until the weeks after reach onset in all groups. These results are particularly interesting given that at the start of the study, infants already demonstrated the ability to bring their hands to their mouths for mouthing and to one another for fingering of the objects attached to their wrists (Rochat, 1989).

What might young infants learn from haptic exploration of objects? First, information processing via haptic exploration provides infants with knowledge about properties of objects, such as hardness, texture, and shape, which cannot be accurately detected by vision alone (Ballesteros et al., 1997; Carello, Grosofsky, Reichel, Solomon, & Turvey, 1989; Case-Smith et al., 1998; Turvey, 1996). This knowledge base allows older infants to discriminate and categorize objects of varying characteristics (Booth, 2006; Bourgeois et al., 2005; Ellis & Oakes, 2006; Palmer, 1989; Ruff, 1984). Yet, infants have weeks to months of experience haptically exploring objects before this discrimination ability emerges (Case-Smith et al., 1998; Rochat, 1989). Besides knowledge about physical properties of objects, what could infants be learning from these early exploration experiences?

Even before infants are able to discriminate physical properties of objects, it has been proposed that they are able to learn what actions those objects afford (E. J. Gibson, 2000). The fact that infants did not use their available movements to explore the Velcroed objects until after the onset of reaching suggests that haptic exploration of objects may in itself be a global affordance that young infants must learn (Bourgeois et al., 2005; J. J. Gibson, 1962, 1966; Millar, 2005; Needham, 2000; Needham et al., 2002). Infants at each visit were shown the exploration objects before trials began, but they disregarded the objects at early visits before they were consistently reaching. For instance, most infants spent less than 10% of the time exploring objects haptically across the first few visits, whereas at the final visit, most infants spent about 40% of the time exploring objects haptically (Figure 4d). Object exploration in this paradigm did not require grasping as objects were Velcroed to infants' wrists. It was as if prereaching infants did not yet realize that they could use their existing arm movements to haptically explore objects. We propose that as young infants begin to reach for, grasp, and transport objects, they learn to match these new action abilities to objects to learn that objects of certain sizes and shapes afford mouthing and fingering (Case-Smith et al., 1998; E. J. Gibson, 2000; Millar, 2005). Then, via early, frequent, active haptic exploration of a range of objects, infants learn more about how to adapt and refine their actions to more specifically match the various properties of objects (Adolph et al., 2000; Ballesteros et al., 1997; Bourgeois et al., 2005; Needham et al., 2006). For instance, infants may later learn that in addition to mouthing and fingering, objects such as rattles and spoons afford sound production and food transport. In effect, learning that some objects can be explored haptically may be the first step in the process of gaining object knowledge required for future object categorization and discrimination.

Task 3: Object and Postural Experience Infants Had Advanced Means – End Behavior

Object experience infants had the most sophisticated means – end behavior, evidenced by means – end behavior during both periods of a visit. Postural experience infants had better developing means–end behavior than social experience infants, evidenced by a greater number of periods of means–end behavior. These results suggest that the developmental origins of means–end behavior may lie in the ongoing motor, perceptual, and cognitive experiences associated with early reaching and object exploration (Bremner, 2000; E. J. Gibson & Pick, 2000; Menard, 2005; Piaget, 1952, 1954; von Helmholtz, 1962). These results highlight the critical role action performance on objects plays in the development of means–end behavior and expand the view that early knowledge development relies primarily on perceptual and cognitive changes (Baillargeon, 1987; Johnson, Slemmer, & Amso, 2004).

Our results suggest that means–end behavior emerges after infants are able to coordinate their motor, perceptual, and affect regulation abilities. Specifically, there were no differences in the individual motor, perceptual, or affect regulation abilities among groups that can fully account for the results. This supports Willatts's (1999) idea that intentional means–end behavior involves the coordination of multiple variables. In our task, infants had to display a greater frequency of toy activations in coordination with looking to the toy while maintaining a positive affect (Lewis et al., 1990; Munakata et al., 2002; Willatts, 1999). Infants' abilities to coordinate these emerging abilities across domains to perform in a novel task are a hallmark of means–end behavior.

Our results support the idea that means–end knowledge is gained in part via active exploration of objects (Bremner, 2000; Piaget, 1952). E. J. Gibson (1997) suggested that infants learn by first performing exploratory activity, then observing the consequences of this activity, and finally selecting from the activities explored. She viewed trial-and-error experience as essential to learning. This type of activity is exactly what infants were required to do to explore the means–end relationship within our paradigm. To learn, infants were required to explore the switches via reaching, observe that movement of one switch resulted in distant toy activity, and then repeatedly select the behavior that resulted in that toy activity by activating that switch more frequently than the other.

Our results provide the first empirical support for the idea that the emergence of means–end behavior is in part dependent on the amount of previous object exploration. Object experience infants had increased opportunities to explore objects from Visit 1, and they demonstrated the most sophisticated means–end behavior. Postural experience infants began reaching and likely had increased opportunities to explore objects from Visit 3 and had more advanced developing means–end behavior than social experience infants. Social experience infants began reaching and likely had increased opportunities to explore objects late in the study from Visit 5 and had the poorest means–end behavior.

The results of this study provide important data on the developmental origins of the means–end behavior so often observed in older infants (Bojczyk & Corbetta, 2004; Goubet, Rochat, Maire-Leblond, & Poss, 2006; Munakata et al., 2002). Using a means–end task such as the one presented here that calls on infants to coordinate abilities they already possess to demonstrate that means–end knowledge may help avoid the performance competency issue encountered by other behavioral means–end studies (Bremner, 2000; Matthews, Ellis, & Nelson, 1996). Means–end paradigms such as ours that capitalize on infants' existing behaviors may help researchers avoid the challenge of comparing violation of expectancy paradigms with behavioral paradigms (Bremner, 2000; Hood, 2001; Munakata et al., 2002; Shinskey, Bogartz, & Poirier, 2000; Sommerville & Woodward, 2005). Future studies can build upon these results to tests for developmental connections between early associative learning and memory paradigms such as Rovee and Rovee's (1969) mobile paradigm and later assessments of object exploration and means–end behavior.

Implications for Developmental Theory and Intervention

The developmental relationships among the three behaviors investigated in this study and the ways the prescribed experiences affected these behaviors support theories of development proposing actions as the foundation for perceptual and cognitive development (Bremner, 2000; E. J. Gibson, 2000; J. J. Gibson, 1979; Piaget, 1952; von Helmholtz, 1962). Our results also have implications for intervention for infants with developmental delay. First, the results support the premise that early intervention can advance the emergence of a behavior (Heathcock et al., 2008; Lobo et al., 2004; Ulrich, Ulrich, Angulo-Kinzler, & Yun, 2001; van Beek, Hopkins, Hoeksma, & Samsom, 1994). Postural and object experiences provided to infants with special needs may be able to advance reaching (Heathcock et al., 2008), just as experience walking supported on a treadmill can advance the emergence of walking in infants with Down syndrome (Ulrich et al., 2001). Although therapeutic interventions often begin in early infancy and frequently rely on techniques attempting to advance postural stability or object interaction, we are not aware of any prospective studies directly relating such controlled movement interventions to advanced behavioral outcomes in infants less than 6 months of age. Second, providing early experiences for one skill may advance development of a range of motor, perceptual, and cognitive abilities (Hack & Taylor, 2000; van Beek et al., 1994). Perhaps similar advancements across developmental domains could be achieved by advancing other early behaviors. For instance, language development may be advanced via experiences to advance early object banging or gesturing (Iverson & Fagan, 2004; Iverson & Goldin-Meadow, 2005) or later problem-solving ability could be advanced via experiences to broaden the ranges of exploratory behaviors young infants perform on various objects (Caruso, 1993). Although most therapists would agree that skill attainment generally advances development across domains, there is little empirical support for this idea in early infancy (Needham, 2000).

In summary, the behavior of infants in this study provides empirical support that (a) the development of novel behaviors is dependent on multiple subsystems and can be similarly advanced by addressing a variety of these subsystems (Bertenthal & von Hofsten, 1998; Hopkins & Ronnqvist, 2002; Thelen, 1990; von Hofsten, 1993); (b) reaching ability can be indirectly advanced via activities to improve postural stability; (c) parent–infant interactions play a critical role in early development and can be altered to advance development (Bakeman et al., 1990; Bruner, 1972; Cintas, 1995; Landry & Chapieski, 1988); (d) reaching, object exploration, and means–end behavior are developmentally related behaviors that can all be advanced via experiences aimed at advancing reaching; and (e) past experiences with active object exploration can facilitate early information processing and the development of early knowledge (Ballesteros et al., 1997; Bremner, 2000; E. J. Gibson & Pick, 2000; Needham, 2000).

References

Adolph, K. E., Eppler, M. A., Marin, L., Weise, I. B., & Wechsler-Clearfield, M. (2000). Exploration in the service of prospective control. *Infant Behavior & Development, 23*, 441–460.

Adolph, K. E., Vereijken, B., & Denny, M. A. (1998). Learning to crawl. *Child Development, 69*, 1299–1312.

Amiel-Tison, C. (1985). Pediatric contribution to the present knowledge on the neurobehavioral status of infants at birth. In J. M. R. Fox (Ed.), *Neonatal cognition: Beyond the blooming buzzing confusion* (pp. 365–380). Hillsdale, NJ: Erlbaum.

Angulo-Kinzler, R. M., & Horn, C. L. (2001). Selection and memory of a lower limb motor-perceptual task in 3-month-old infants. *Infant Behavior & Development, 24*, 239–257.

Baillargeon, R. (1987). Object permanence in 3½- and 4½-month-old infants. *Developmental Psychology, 23*, 655–664.

Bakeman, R., Adamson, L. B., Konner, M., & Barr, R. G. (1990). Kung infancy: The social context of object exploration. *Child Development, 61*, 794–809.

Ballesteros, S., Manga, D., & Reales, J. M. (1997). Haptic discrimination of bilateral symmetry in 2-dimensional and 3-dimensional unfamiliar displays. *Perception & Psychophysics, 59*, 37–50.

Bertenthal, B., & von Hofsten, C. (1998). Eye, head and trunk control: The foundation for manual development. *Neuroscience and Biobehavioral Reviews, 22*, 515–520.

Berthier, N. E., Rosenstein, M. T., & Barto, A. G. (2005). Approximate optimal control as a model for motor learning. *Psychological Review, 112*, 329–346.

Bhat, A. N., & Galloway, J. C. (2006). Toy-oriented changes during early arm movements: Hand kinematics. *Infant Behavior & Development, 29*, 358–372.

Bhat, A., Heathcock, J., & Galloway, J. C. (2005). Toy-oriented changes in hand and joint kinematics during the emergence of purposeful reaching. *Infant Behavior & Development, 28*, 445–464.

Bojczyk, K. E., & Corbetta, D. (2004). Object retrieval in the 1st year of life: Learning effects of task exposure and box transparency. *Developmental Psychology, 40*, 54–66.

Booth, A. E. (2006). Object function and categorization in infancy: Two mechanisms of facilitation. *Infancy, 10*, 145–169.

Bourgeois, K. S., Khawar, A. W., Neal, S. A., & Lockman, J. J. (2005). Infant manual exploration of objects, surfaces, and their interrelations. *Infancy, 8*, 233–252.

Bremner, J. G. (2000). Developmental relationships between perception and action in infancy. *Infant Behavior & Development, 23*, 567–582.

Bril, B., & Sabatier C. (1986). The cultural context of motor development: Postural manipulations in the daily life of Bambara babies (Mali). *International Journal of Behavioral Development, 9*, 439–453.

Bruner, J. (1972). The nature and uses of immaturity. *American Psychologist, 27*, 687–708.

Campos, J. J., Anderson, D. I., Barbu-Roth, M. A., Hubbard, E. M., Hertenstein, M. J., & Witherington, D. (2000). Travel broadens the mind. *Infancy, 1*, 149–219.

Carello, C., Grosofsky, A., Reichel, F. D., Solomon, H. Y., & Turvey, M. T. (1989). Visually perceiving what is reachable. *Ecological Psychology, 1*, 27–54.

Caruso, D. A. (1993). Dimensions of quality in infants' exploratory behavior: Relationships to problem-solving ability. *Infant Behavior & Development, 16*, 441–454.

Carvalho, R. P., Tudella, E., & Savelsbergh, G. J. P. (2007). Spatio-temporal parameters in infant's reaching movements are influenced by body orientation. *Infant Behavior & Development, 30*, 26–35.

Case-Smith, J., Bigsby, R., & Clutter, J. (1998). Perceptual-motor coupling in the development of grasp. *American Journal of Occupational Therapy, 52*, 102–110.

Cintas, H. (1995). Cross-cultural similarities and differences in development and the impact of parental expectations on motor behavior. *Pediatric Physical Therapy, 7,* 103–111.

Corbetta, D., & Bojczyk, K. E. (2002). Infants return to two-handed reaching when they are learning to walk. *Journal of Motor Behavior, 34,* 83–95.

Corbetta, D., Thelen, E., & Johnson, K. (2000). Motor constraints on the development of perception-action matching in infant reaching. *Infant Behavior & Development, 23,* 351–374.

Davis, B. E., Moon, R. Y., Sachs, H. C., & Ottolini, M. C. (1998). Effects of sleep position on infant motor development. *Pediatrics, 102,* 1135–1140.

Ellis, A. E., & Oakes, L. M. (2006). Infants flexibly use different dimensions to categorize objects. *Developmental Psychology, 42,* 1000–1011.

Eppler, M. A. (1995). Development of manipulatory skills and the deployment of attention. *Infant Behavior & Development, 18,* 391–405.

Fagard, J. (2000). Linked proximal and distal changes in the reaching behavior of 5- to 12-month-old human infants grasping objects of different sizes. *Infant Behavior & Development, 23,* 317–329.

Fallang, B., Saugstad, O. D., & Hadders-Algra, M. (2000). Goal directed reaching and postural control in supine position in healthy infants. *Behavioural Brain Research, 115,* 9–18.

Fogel, A. (1997). Information, creativity, and culture. In C. D.-R. P. Zukow-Goldring (Ed.), *Evolving explanations of development.* Washington, DC: American Psychological Association.

Fogel, A., Messinger, D. S., Dickson, K. L., & Hsu, H. (1999). Posture and gaze in early mother-infant communication: Synchronization of developmental trajectories. *Developmental Science, 2,* 325–332.

Fontaine, R., & le Bonniec, G. P. (1988). Postural evolution and integration of the prehensile gesture in children aged 4 to 10 months. *British Journal of Developmental Psychology, 6,* 223–233.

Funk, M. S. (2002). Problem solving skills in young yellow-crowned parakeets (cyanoramphus auriceps). *Animal Cognition, 5,* 167–176.

Gibson, E. J. (1988). Exploratory behavior in the development of perceiving, acting, and the acquiring of knowledge. *Annual Review of Psychology, 39,* 1–41.

Gibson, E. J. (1997). An ecological psychologist's prolegomena for perceptual development: A functional approach. In C. Dent-Read & P. Zukow-Goldring (Eds.), *Evolving explanations of development* (pp. 23–54). Washington, DC: American Psychological Association.

Gibson, E. J. (2000). Perceptual learning in development: Some basic concepts. *Ecological Psychology, 12,* 295–302.

Gibson, E. J., & Pick, A. D. (2000). *An ecological approach to perceptual learning and development.* Oxford, UK: Oxford University Press.

Gibson, J. J. (1962). Observations on active touch. *Psychological Review, 69,* 477–491.

Gibson, J. J. (1966). *The senses considered as perceptual systems.* Boston: Houghton Mifflin.

Gibson, J. J. (1979). *The ecological approach to visual perception.* Boston: Houghton Mifflin.

Goubet, N., Rochat, P., Maire-Leblond, C., & Poss, S. (2006). Learning from others in 9-18-month-old infants. *Infant and Child Development, 15,* 161–177.

Gowen, J. W., Goldman, B. D., Johnson-Martin, N., & Hussey, B. (1989). Object play and exploration of handicapped and nonhandicapped infants. *Journal of Applied Developmental Psychology, 10,* 53–72.

Grenier, A. (1981). Liberated-motricity by holding the head during the 1st weeks of life. *Archives Francaises De Pediatrie, 38,* 557–561.

Hack, M., & Taylor, H. G. (2000). Perinatal brain injury in preterm infants and later neurobehavioral function. *Journal of the American Medical Association, 284,* 1973–1974.

Hadders-Algra, M., Brogren, E., & Forssberg, H. (1996). Ontogeny of postural adjustments during sitting in infancy: Variation, selection and modulation. *Journal of Physiology, 493,* 273–288.

Harris, F. C., & Lahey, B. B. (1978). A method for combining occurrence and nonoccurrence interobserver agreement scores. *Journal of Applied Behavioral Analysis, 11,* 523–527.

Heathcock, J. C., Lobo, M. A., & Galloway, J. C. (2008). Movement training advances the emergence of reaching in infants born preterm at < 33 weeks gestational age. *Physical Therapy, 88*(3), 1–13.

Hofstadter, M., & Reznick, J. S. (1996). Response modality affects human infant delayed-response performance. *Child Development, 67,* 646–658.

Hood, B. (2001). Guest editorial—When do infants know about objects? *Perception, 30,* 1281–1284.

Hopkins, B., & Ronnqvist, L. (2002). Facilitating postural control: Effects on the reaching behavior of 6-month-old infants. *Developmental Psychobiology, 40,* 168–182.

Hopkins, B., & Westra, T. (1988). Maternal handling and motor development: An intracultural study. *Genetic, Social, and General Psychology Monographs, 114,* 379–408.

Hopkins, B., & Westra, T. (1989). Maternal expectations of their infants' development: Some cultural differences. *Developmental Medicine and Child Neurology, 31,* 384–390.

Hopkins, B., & Westra, T. (1990). Motor development, maternal expectations, and the role of handling. *Infant Behavior and Development, 13,* 117–122.

Iverson, J. M., & Fagan, M. K. (2004). Infant vocal-motor coordination: Precursor to the gesture-speech system? *Child Development, 75,* 1053–1066.

Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science, 16,* 367–371.

Izard, C. E., Dougherty, L. M., & Hembree, E. A. (1989). *A System for Identifying Affect Expressions by Holistic Judgments (AFFEX).* Newark: University of Delaware.

Joh, A. S., & Adolph, K. E. (2006). Learning from falling. *Child Development, 77,* 89–102.

Johnson, S. P., Slemmer, J. A., & Amso, D. (2004). Where infants look determines how they see: Eye movements and object perception performance in 3-month-olds. *Infancy, 6*, 185–201.

Konczak, J., & Dichgans, J. (1997). The development toward stereotypic arm kinematics during reaching in the first 3 years of life. *Experimental Brain Research, 117*, 346–354.

Landry, S. H., & Chapieski, M. L. (1988). Visual attention during toy exploration in preterm infants: Effects of medical risk and maternal interactions. *Infant Behavior & Development, 11*, 187–204.

Lee, M., Liu, Y., & Newell, K. M. (2006). Longitudinal expressions of infant's prehension as a function of object properties. *Infant Behavior & Development, 29*, 481–493.

Lewis, M., Alessandri, S. M., & Sullivan, M. W. (1990). Violation of expectancy, loss of control, and anger expressions in young infants. *Developmental Psychology, 26*, 745–751.

Lewis, M., Hitchcock, D. F. A., & Sullivan, M. W. (2004). Physiological and emotional reactivity to learning and frustration. *Infancy, 6*, 121–143.

Lobo, M. A., Galloway, J. C., & Savelsbergh, G. J. P. (2004). General and task-related experiences affect early object interaction. *Child Development, 75*, 1268–1281.

Matthews, A., Ellis, A. E., & Nelson, C. A. (1996). Development of preterm and full-term infant ability on AB, recall memory, transparent barrier detour, and means-end tasks. *Child Development, 67*, 2658–2676.

Menard, K. R. (2005). *Means-end search for hidden objects by 6.5-month-old infants: Examination of an experiential limitation hypothesis*. Doctoral thesis, Univerisity of Waterloo.

Mildred, J., Beard, K., Dallwitz, A., & Unwin, J. (1995). Play position is influenced by knowledge of SIDS sleep position recommendations. *Journal of Pediatric Child Health, 31*, 499–502.

Millar, S. (2005). Network models for haptic perception. *Infant Behavior & Development, 28*, 250–265.

Munakata, Y., Bauer, D., Stackhouse, T., Landgraf, L., & Huddleston, J. (2002). Rich interpretation vs. deflationary accounts in cognitive development: The case of means-end skills in 7-month-old infants. *Cognition, 83*, B43–B53.

Needham, A. (2000). Improvements in object exploration skills may facilitate the development of object segregation in early infancy. *Journal of Cognition and Development, 1*, 131–156.

Needham, A., Barrett, T., & Peterman, K. (2002). A pick-me-up for infants' exploratory skills: Early simulated experiences reaching for objects using 'sticky mittens' enhances young infants' object exploration skills. *Infant Behavior & Development, 25*, 279–295.

Needham, A., Cantlon, J. F., & Holley, S. M. O. (2006). Infants' use of category knowledge and object attributes when segregating objects at 8.5 months of age. *Cognitive Psychology, 53*, 345–360.

Palmer, C. F. (1989). The discriminating nature of infants exploratory actions. *Developmental Psychology, 25*, 885–893.

Piaget, J. (1952). *The origins of intelligence in children*. New York: Norton.

Piaget, J. (1954). *The construction of reality in the child*. New York: Basic Books.

Piper, M. C., & Darrah, J. (1994). *Motor assessment of the developing infant*. Philadelphia: W.B. Saunders.

Piper, M. C., Pinnell, L. E., Darrah, J., Maguire, T., & Byrne, P. J. (1992). Construction and validation of the Alberta Infant Motor Scale (AIMS). *Canadian Journal of Public Health, 83*(2), 46–50.

Prechtl, H. F. R. (1974). The behavioral states of the newborn infant (a review). *Brain Research, 76*, 185–212.

Provine, R. R., & Westerman, J. A. (1979). Crossing the midline: Limits of early eye-hand behavior. *Child Development, 50*, 437–441.

Reed, E. S., & Bril, B. (1996). The primacy of action in development. In M. T. Turvey (Ed.), *Dexterity and its development* (pp. 431–451). Mahwah, NJ: Erlbaum.

Rochat, P. (1989). Object manipulation and exploration in 2-month-old to 5-month- old infants. *Developmental Psychology, 25*, 871–884.

Rochat, P. (1992). Self-sitting and reaching in 5- to 8-month-old infants: The impact of posture and its development on early eye-hand coordination. *Journal of Motor Behavior, 24*, 210–220.

Rochat, P., & Goubet, N. (1995). Development of sitting and reaching in 5- to 6-month-old infants. *Infant Behavior & Development, 18*, 53–68.

Rovee, C. K., & Rovee, D. T. (1969). Conjugate reinforcement of infant exploratory behavior. *Journal of Experimental Child Psychology, 8*, 33–39.

Ruff, H. A. (1984). Infants' manipulative exploration of objects: Effects of age and object characteristics. *Developmental Psychology, 20*, 9–20.

Savelsbergh, G. J. P., & van der Kamp, J. (1993). The coordination of infant's reaching, grasping, catching and posture: A natural physical approach. *Advances in Psychology, 97*, 289–317.

Savelsbergh, G. J. P., & van der Kamp, J. (1994). The effect of body orientation to gravity on early infant reaching. *Journal of Experimental Child Psychology, 58*, 510–528.

Shinskey, J. L., Bogartz, R. S., & Poirier, C. R. (2000). The effects of graded occlusion on manual search and visual attention in 5- to 8- month old infants. *Infancy, 1*, 323–346.

Shinskey, J. L., & Munakata, Y. (2001). Detecting transparent barriers: Clear evidence against the means-end deficit account of search failures. *Infancy, 2*, 395–404.

Shumway-Cook, A., & Woollacott, M. H. (1995). *Motor control: Theory and practical applications*. Philadelphia: Williams & Wilkins.

Snow, C. W. (1989). *Infant development*. Upper Saddle River, NJ: Prentice Hall.

Sommerville, J. A., & Woodward, A. L. (2005a). Infants' sensitivity to the causal features of means-end support sequences in action and perception. *Infancy, 8*, 119–145.

Sommerville, J. A., & Woodward, A. L. (2005b). Pulling out the intentional structure of action of action: The relation between action processing and action production infancy. *Cognition*, *95*, 1 – 30.

Spencer, J. P., & Thelen, E. (2000). Spatially specific changes in infants' muscle coactivity as they learn to reach. *Infancy*, *1*, 275 – 302.

Thelen, E. (1990). Coupling perception and action in the development of skill: A dynamic approach. In H. Bloch & B. I Berthenthal (Eds.), *Sensory-motor organizations and development in infancy and early childhood*. Boston: Kluwer Academic.

Thelen, E., Corbetta, D., Kamm, K., Spencer, J. P., Schneider, K., & Zernicke, R. F. (1993). The transition to reaching: Mapping intention and intrinsic dynamics. *Child Development*, *64*, 1058 – 1098.

Thelen, E., & Spencer, J. P. (1998). Postural control during reaching in young infants: A dynamic systems approach. *Neuroscience and Biobehavioral Reviews*, *22*, 507 – 514.

Turvey, M. T. (1996). Dynamic touch. *American Psychologist*, *51*, 1134 – 1152.

Ulrich, D. A., Ulrich, B. D., Angulo-Kinzler, R. M., & Yun, J. (2001). Treadmill training of infants with Down syndrome: Evidence-based developmental outcomes. *Pediatrics*, *108*, U42 – U48.

van Beek, Y., Hopkins, B., Hoeksma, J. B., & Samsom, J. F. (1994). Prematurity, posture and the development of looking behavior during early communication. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *35*, 1093 – 1107.

van Hof, P., van der Kamp, J., & Savelsbergh, G. J. P. (2002). The relation of unimanual and bimanual reaching to crossing the midline. *Child Development*, *73*, 1353 – 1362.

Vereijken, B., & Thelen, E. (1997). Training infant treadmill stepping: The role of individual pattern stability. *Developmental Psychobiology*, *30*, 89 – 102.

von Helmholtz, H. (1962). *Handbook of physiological optics* (Rev. ed.). New York: Optical Society of America.

von Hofsten, C. (1991). Structuring of early reaching movements: A longitudinal study. *Journal of Motor Behavior*, *23*, 280 – 292.

von Hofsten, C. (1993). Prospective control: A basic aspect of action development. *Human Development*, *36*, 253 – 270.

von Hofsten, C. (1997). On the early development of predictive abilities. In C. Dent-Read & P. Zukow-Goldring (Eds.), *Evolving explanations of development* (pp. 163 – 194). Washington, DC: American Psychological Association.

Wijnroks, L., & van Veldhoven, N. (2003). Individual differences in postural control and cognitive development in preterm infants. *Infant Behavior & Development*, *26*, 14 – 26.

Willatts, P. (1999). Development of means-end behavior in young infants: Pulling a support to retrieve a distant object. *Developmental Psychology*, *35*, 651 – 667.

Wimmers, R. H., Savelsbergh, G. J. P., Beek, P. J., & Hopkins, B. (1998). Evidence for a phase transition in the early development of prehension. *Developmental Psychobiology*, *32*, 235 – 248.

Wimmers, R. H., Savelsbergh, G. J. P., van der Kamp, J., & Hartelman, P. (1998). A developmental transition in prehension modeled as a cusp catastrophe. *Developmental Psychobiology*, *32*, 23 – 35.

Witherington, D. C., von Hofsten, C., Rosander, K., Robinette, A., Woollacott, M. H., & Bertenthal, B. I. (2002). The development of anticipatory postural adjustments in infancy. *Infancy*, *3*, 495 – 517.

Zelazo, N. A., Zelazo, P. R., Cohen, K. M., & Zelazo, P. D. (1993). Specificity of practice effects on elementary neuromotor patterns. *Developmental Psychology*, *29*, 686 – 691.

Child Development, November/December 2008, Volume 79, Number 6, Pages 1891–1906

A Process Model of Attachment–Friend Linkages: Hostile Attribution Biases, Language Ability, and Mother–Child Affective Mutuality as Intervening Mechanisms

Nancy L. McElwain
University of Illinois at Urbana–Champaign

Cathryn Booth-LaForce
University of Washington

Jennifer E. Lansford
Duke University

Xiaoying Wu and W. Justin Dyer
University of Illinois at Urbana–Champaign

This study identified mechanisms through which child–mother attachment security at 36 months was associated with mother- and teacher-reported friendship quality at 3rd grade. Data from a subsample of 1,071 children (536 boys) participating in the National Institute of Child Health and Human Development Study of Early Child Care and Youth Development were used. Separate structural equation models were tested for mother and teacher reports of peer functioning. For both models, the total indirect effect between attachment security and friendship quality was significant. Tests of specific indirect effects indicated that attachment security was associated with friendship quality via greater mother–child affective mutuality and better language ability at 54 months and fewer hostile attributions (teacher model only) and greater peer competence at first grade. The findings highlight interpersonal and intrapersonal mechanisms of attachment–friend linkages.

Establishing and maintaining close friendships is a central developmental task of middle childhood, yet the quality of children's friendships varies widely (Hartup, 1996). Although not the only contributor to friendship quality, the intimate, dyadic nature of attachment relationships is posited to be especially germane to children's close relationships, such as friendships, compared with more general indices of social adjustment (see Belsky & Cassidy, 1994; Sroufe & Fleeson, 1986; Thompson, 1998). In line with this view, child–mother attachment security has been related in theoretically expected ways to children's friendship quality (e.g., Elicker, Englund, & Sroufe, 1992; McElwain, Cox, Burchinal, & Macfie, 2003; McElwain & Volling, 2004; Youngblade & Belsky, 1992), and in a meta-analysis of attachment–peer associations, Schneider, Atkinson, and Tardiff (2001) reported a larger effect size for studies of children's close friendships versus studies of children's interactions with unfamiliar peers (also see Lieberman, Doyle, & Markiewicz, 1999). We aimed to further this line of inquiry by assessing potential intervening mechanisms of attachment–friend linkages.

Our investigation was guided by attachment theory, in which Bowlby (1969/1982) proposed that children continuously construct "working models" of self, other, and relationships through repeated interactions with caregivers. The internal working model, in turn, mediates associations between early attachment relationships and later outcomes through the formation of expectations or rules for behavior (see Bretherton & Munholland, 1999) and the filtering of the individual's attention and memory in ways that are consistent with his or her expectations for interpersonal interactions (see Main, 1995; Main, Kaplan, & Cassidy, 1985). Notably, patterns of verbal and nonverbal communication between the caregiver and the child, especially about emotions, are considered central to the development and maintenance of working models of attachment (Bowlby, 1969/1982, 1988; Bretherton, 1990; Kobak & Duemmler, 1994; Main et al., 1985). From birth, caregivers respond to their infant's vocal and nonvocal cues, and these caregiver–infant exchanges—repeated over time—lay the

---

The National Institute of Child Health and Human Development (NICHD) Study of Early Child Care and Youth Development (SECCYD) is directed by a Steering Committee and supported by NICHD through a cooperative agreement (U10) that calls for a scientific collaboration between the grantees and the NICHD staff. We wish to express our appreciation to the principal investigators, site coordinators, and participants of the NICHD SECCYD.

Correspondence concerning this article should be addressed to Nancy L. McElwain, Department of Human and Community Development, University of Illinois at Urbana–Champaign, 904 West Nevada Street, Urbana, IL 61801. Electronic mail may be sent to mcelwn@illinois.edu.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7906-0024

foundation for nascent working models of self and other (Ainsworth, Blehar, Waters, & Wall, 1978). During the preschool years, the attachment relationship is transformed by the child's burgeoning language and cognitive skills, and Thompson (2000) proposed that it is during this period of representational advance that working models of attachment may be especially influential in predicting child functioning. In light of these propositions, we assessed children's hostile attribution biases and language ability as intervening mechanisms linking child–mother attachment security at 36 months and children's friendship quality at Grade 3. Furthermore, open emotional communication between mothers and children, which we refer to here as affective mutuality, was examined as an interpersonal process through which child–mother attachment security would be associated with intrapersonal processes (i.e., hostile attributions, language ability) and ultimately friendship quality.

Hostile Attribution Biases

Dodge and colleagues' (Crick & Dodge, 1994; Dodge, Pettit, McClaskey, & Brown, 1986) model of social information processing (SIP) includes several steps (e.g., encoding and interpreting social cues, generating and evaluating potential responses) that are expected to influence an individual's response to a social stimulus. We focus here on attributions of hostile intent when interpreting social cues because this SIP construct, in particular, parallels the notion of the internal working model as an interpretative filter (e.g., Main, 1995). The utility of hostile attributions as an intervening mechanism of attachment–friend linkages is also suggested by studies linking hostile attributions to a host of peer-related outcomes, including aggression (see Orobio de Castro, Veerman, Koops, Bosch, & Monshouwer, 2002), peer rejection (Feldman & Dodge, 1987), friendship quality, and social competence (Webster-Stratton & Lindsay, 1999).

Importantly, the SIP model highlights "latent mental structures" as the individual's scripts, schemata or cognitive heuristics of previous experiences that have been integrated into long-term memory and that guide processing of social stimuli (Crick & Dodge, 1994; Dodge & Schwartz, 1997). Reliance on latent mental structures, which is especially likely when social stimuli are ambiguous, allows for more efficient, schema-consistent processing of environmental cues but may also lead to errors or distortions (Baldwin, 1992; Crick & Dodge, 1994). The SIP notion of latent mental structures echoes Bretherton and Munholland's (1999) discussion of the internal working model as event schemas or scripts. Thus, the child's internal working model (or latent mental structures), which emerges from repeated interactions with attachment figures, is expected to guide online interpretation of social stimuli. For instance, guided by a working model of the self as worthy of care and others as supportive and trustworthy, a child with a secure attachment history may make benign attributions about negative interpersonal events in which the intent of the other is ambiguous. In contrast, a child with an insecure attachment history, and corresponding working model of the self as unworthy and others as unresponsive, may be more likely to attribute hostile intentions to others. In other words, a child who has received repeated rejection and hostility from attachment figures may come to expect such treatment from others, even when situational cues do not merit such expectations.

Of the few studies investigating associations between attachment and hostile attribution biases, evidence has been mixed. Suess, Grossmann, and Sroufe (1992) reported that an insecure–avoidant infant–mother attachment was associated with attributions of negative peer intent at age 5 (but see Cassidy, Kirsh, Scolton, & Parke, 1996; Ziv, Oppenheim, & Sagi-Schwartz, 2004, for null findings), and Cassidy et al. (1996) found that attachment security during the early school years was concurrently related to more positive attributions of peer intent. Little is known about preschool attachment security as a predictor of hostile attribution biases. As highlighted by Crittenden (1990), working models move from reliance on procedural memories during infancy to integration of semantic memories during late toddlerhood. Unlike procedural memories, semantic memories are conscious, rely on language for encoding and recall, and are likely to be elicited when information is ambiguous. Because these developmental changes result in more complex and language-mediated working models, attachment security during the early preschool period may be particularly important for setting in motion attributional biases that, in turn, impact later functioning (see also Thompson, 2000).

Only one study, to our knowledge, has tested SIP biases as a mechanism of attachment–friend associations. Cassidy et al. (1996, see study 2) examined a composite of peer representations measured via Dodge and Frame's (1982) SIP task as a mediator of concurrent associations between child–mother attachment security and number of reciprocal friendships in the kindergarten or first-grade classroom. No evidence of mediation was found. Rather, attachment security and peer representations were each significant predictors of the outcome, but only when entered first in the model. The small sample and concurrent measurement of the predictor, mediator, and outcome variables may have hindered detection of mediation in the Cassidy et al. study. The current study addressed these limitations.

Language Ability

In highlighting the role of communicative processes for developing attachment relationships, attachment theorists have considered the bidirectional influence of attachment and language. The child's emerging language skills are expected to transform his or her attachment relationships, and attachment relationships may also forecast the child's competence in acquiring and using language (Bowlby, 1969/1982; Bretherton, 1990; Main et al., 1985). This latter possibility is further suggested by research underscoring the social foundations of language acquisition (e.g., see Tomasello, 1992) and the notion that, from an evolutionary standpoint, language may be rooted in vocal behaviors aimed at fulfilling relational goals (Locke, 2001). Although few researchers have investigated attachment security as an antecedent of language competence, van IJzendoorn, Dijkstra, and Bus (1995) conducted a meta-analytic review of seven studies and reported a substantial association (r = .28) between infant–mother attachment security and greater language ability. Moreover, examining data from the National Institute of Child Health and Human Development (NICHD) Study of Early Child Care and Youth Development (SECCYD), Belsky and Fearon (2002) found that secure infant–mother attachment was associated with language comprehension at 36 months, controlling for contextual risk. Language and attachment researchers alike have suggested that the child's interactions with a sensitive, responsive caregiver may account, to some extent, for the better language abilities of children with a history of secure attachment (e.g., Cohen, 2001; Fish & Pinkerman, 2003; van IJzendoorn et al., 1995).

Language also underpins key processes of childhood friendship formation and maintenance, including establishing a common ground, constructing joint plans, coordinating play, and negotiating resolutions to conflict (Gottman, 1983; Howes, Droege, & Matheson, 1994; Parker & Gottman, 1989). Not surprisingly, then, language difficulties have been related to lower social competence and poorer quality friendships (see Farmer, 2006), whereas higher levels of expressive and receptive language abilities have been associated with more positive, cooperative peer interactions (e.g., Mendez, Fantuzzo, & Cicchetti, 2002) and more socially connected interactions with friends (e.g., Hebert-Myers, Guttentag, Swank, Smith, & Landry, 2006). Given the importance of verbal communication for children's relationships with attachment figures and friends, we investigated child language ability as a second intervening mechanism.

Mother–Child Affective Mutuality

Affective mutuality (i.e., the balanced, free-flowing verbal and nonverbal exchange of emotions) is a core characteristic of secure attachment relationships. With the advent of language, open verbal communication about emotions has the potential to foster the child's feelings of security, exploration of his or her inner world, and effectiveness in coping with negative emotions (Bowlby, 1988; Bretherton, 1990; Kobak & Duemmler, 1994; Main et al., 1985). In contrast, exchanges in which the attachment figure censures, negates, or distorts the child's emotional experience are expected to have negative consequences for child well-being due to conflicting working models (e.g., what is experienced vs. what is said), impoverished exploration of attachment-related thoughts and feelings, and difficulty integrating and updating working models of self and other. Accordingly, infant–mother attachment security has been associated longitudinally with more emotionally open and coherent mother–child communication, whereas a history of insecure attachment has been associated with communication that is rigid, restricted, and less coherent (e.g., Etzion-Carasso & Oppenheim, 2000; Main et al., 1985). Likewise, attachment security assessed during the preschool years has been related to more frequent, elaborative discourse about negative emotions during mother–child talk about past events (Farrar, Fasig, & Welch-Ross, 1997; Laible, 2004). Open emotional exchange between mothers and children, in turn, has been related to children's higher levels of language and concept development (Pianta, Nimetz, & Bennett, 1997), greater social competence (Pianta et al., 1997), and more constructive coping strategies (Gentzler, Contreras-Grau, Kerns, & Weimer, 2005). In sum, mother–child affective mutuality may be an important interpersonal process through which attachment security is associated with intrapersonal competencies.

The Current Study

In assessing hostile attributions, language ability, and affective mutuality as mechanisms of attachment–friend linkages, we used data from the NICHD SECCYD. Model A in Figure 1 presents the hypothesized process model. This model posits chains of indirect pathways from child–mother attachment

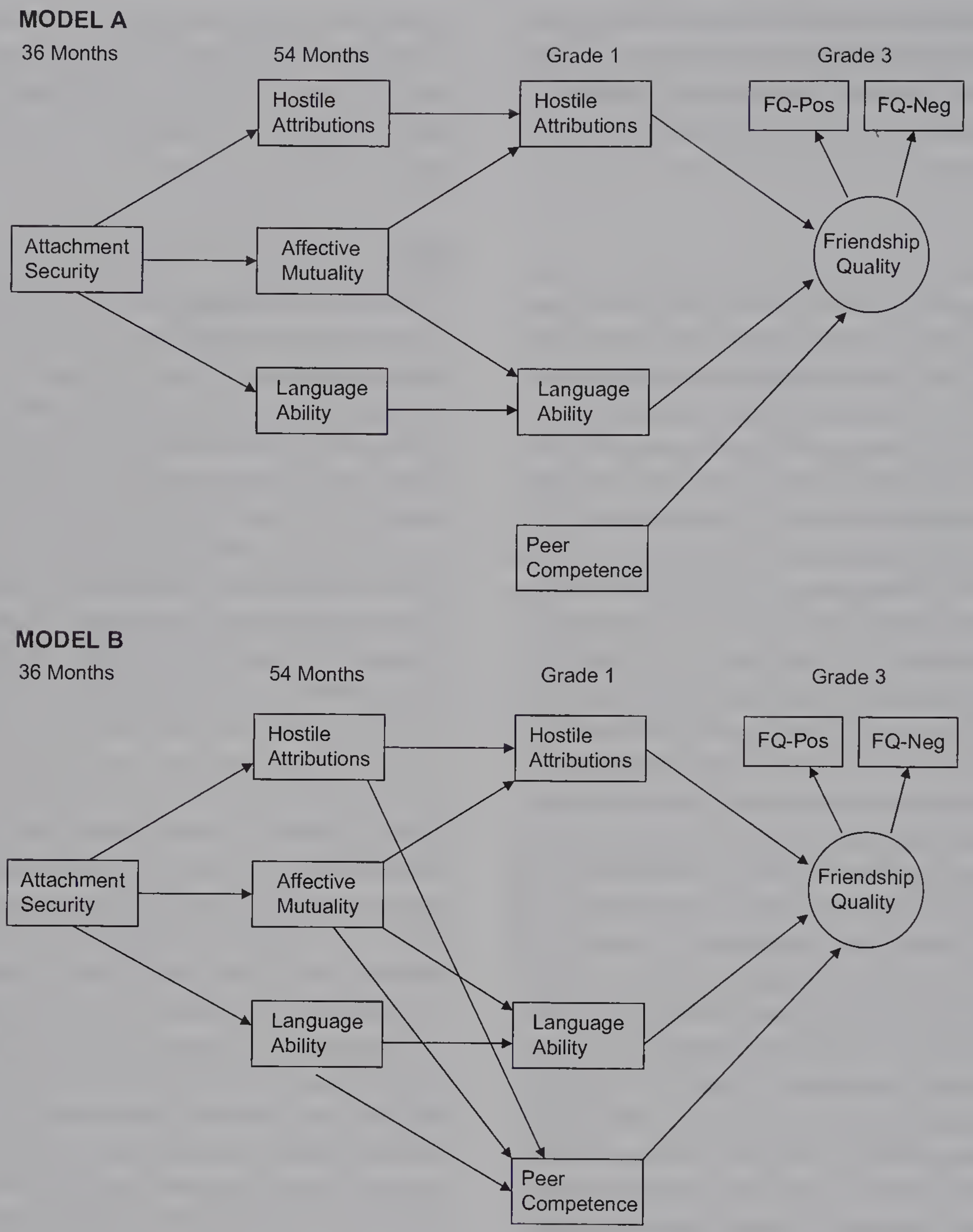

MODEL A
36 Months
54 Months
Grade 1
Grade 3
Hostile Attributions
Hostile Attributions
FQ-Pos
FQ-Neg
Attachment Security
Affective Mutuality
Friendship Quality
Language Ability
Language Ability
Peer Competence
MODEL B
36 Months
54 Months
Grade 1
Grade 3
Hostile Attributions
Hostile Attributions
FQ-Pos
FQ-Neg
Attachment Security
Affective Mutuality
Friendship Quality
Language Ability
Language Ability
Peer Competence

Figure 1. Model A: Children's hostile attributions and language ability (54 months and Grade 1) and mother–child affective mutuality (54 months) as intervening mechanisms between child–mother attachment at 36 months and friendship quality at Grade 3. Model B: Peer competence at Grade 1 as an additional intervening mechanism.
Note. FQ-Pos = friendship quality, positive interaction subscale; FQ-Neg = friendship quality, negative interaction subscale.

security and affective mutuality to hostile attributions and language ability and from hostile attributions and language ability to friendship quality, which is conceptualized as high levels of positive features and low levels of negative features (Berndt, 2002). We expected that mother–child affective mutuality, as an interpersonal process, would influence friendship quality via intrapersonal processes (i.e., hostile attributions, language ability). Moreover, a secondary aim was to assess the degree to which 36-month attachment security was related to mechanisms in Grade 1 via mother–child affective mutuality at 54 months.

The model is guided by strong theoretical assumptions regarding the causal nature and temporal ordering of constructs. In real time, the child–mother relationship is expected to precede and shape the child's attributional biases and language abilities, which, in turn, are expected to precede and shape the child's behavior and relationships with others. Furthermore, the longitudinal design and repeated assessments of constructs permit an analysis of change over time, and this "replicative strategy" (Hoyle & Robinson, 2004) is considered optimal when testing intervening mechanisms in nonexperimental studies (e.g., see Cole & Maxwell, 2003).

We also examined alternative indirect pathways in which peer competence in Grade 1 was part of the chain of intervening mechanisms (see Model B in Figure 1). Fewer hostile attributions (see Crick & Dodge, 1994) and better language skills (see Farmer, 2006) have been associated with social adjustment, and these proposed mechanisms may be associated with friendship quality via peer competence. Furthermore, during interactions with a sensitive, emotionally open caregiver, children may learn skills such as reciprocity and empathy (Elicker et al., 1992; Lieberman, 1977), which foster peer competence and, in turn, high-quality friendships. These alternative indirect paths were added to Model A to examine whether their addition significantly improved model fit.

Mother and teacher reports were used to assess peer competence in Grade 1 and friendship quality in Grade 3. Mother–teacher agreement on children's social competence and peer functioning tends to be modest (see Renk & Phares, 2004), and discrepancies may be due, in part, to differences in child behavior and/or peer partners across home and school contexts. Hence, we tested the process model separately for mother and teacher reports, and we explored in follow-up analyses whether the strength of paths predicting peer competence and friendship quality differed by reporter.

Finally, child gender was examined as a moderator of the process model. "Moderated mediation" occurs when an intervening process varies as a function of a third variable (Muller, Judd, & Yzerbyt, 2005). We expected attachment security to relate to hostile attributions and language ability similarly for girls and boys, yet associations between these variables and friendship quality may differ by gender. Because girls tend to be more interpersonally oriented (e.g., Leaper, 1991), the model may be stronger for girls. On the other hand, research on hostile attributions has largely focused on boys, and it is unknown whether associations between hostile attributions and peer outcomes are stronger for boys (Orobio de Castro et al., 2002).

Method

Participants

Participants were a subsample of 1,071 families drawn from the larger sample of 1,364 families participating in the NICHD SECCYD. Families were recruited from hospitals located in or near 10 sites across the United States and were screened on numerous criteria (e.g., mothers aged 18 years and younger were excluded from the study; all infants were healthy; see NICHD Early Child Care Research Network [ECCRN], 1997, 1999, for further details). The subsample examined in the current report had complete data on mother and/or teacher reports of friendship quality in third grade.

The demographic characteristics for the subsample are presented in Table 1. On average, mothers were 28.49 years of age and had received 14.44 years of education, as reported by mothers when study children were 1 month of age. The income-to-needs ratio (averaged across data collected at 6, 15, 24, and 36 months) was 3.72, indicating that overall family income was, on average, 3.72 times that of the poverty level. In assessing maternal partner status, mothers were asked at 1, 6, 15, 24, and 36 months whether they were currently living with a partner or husband. For 76% of the families, a partner or spouse lived in the home at all time points. Fifty percent of the study children were male. With respect to child ethnicity, 78% were European American non-Hispanic, 11% were African American non-Hispanic, 6% were Hispanic, and 5% were another race or more than one race.

Overview of Procedure

Child–mother attachment was assessed using a modified Strange Situation procedure at 36 months, and mother–child affective mutuality was assessed from a series of mother–child interactive tasks at 54 months. For coding purposes, all videotapes for a given assessment were shipped to a central site, and different coding teams (who were blind to participant characteristics and other study results) assessed child–mother attachment security and affective mutuality. Child interviews were conducted at 54 months and Grade 1 to assess hostile attributions and language ability. Mothers and teachers completed questionnaires assessing children's peer competence at Grade 1 and friendship quality at Grade 3.

Mother–Child Relationship

Attachment security. At 36 months, the modified 17-min Strange Situation procedure developed by Cassidy and Marvin (1992) was used to assess child–mother attachment security. This procedure took place in a laboratory playroom and consisted of five episodes: (a) a 3-min warm-up period, (b) a 3-min separation from mother, (c) a 3-min reunion with mother, (d) a 5-min separation, and (e) a second 3-min reunion. Like the Ainsworth et al. (1978) system for coding infant attachment, the Cassidy and Marvin system emphasizes the child's behavior during the reunion episodes, and children were classified into

Table 1
Descriptive Statistics on Demographic and Study Measures for Cases Included Versus Excluded

Measures	Cases included		Cases excluded	
Demographic measures	*N*	*M* (*SD*)	*N*	*M* (*SD*)
Maternal age	1,071	28.49 (5.58)	293	26.74 (5.62)
Maternal education	1,071	14.44 (2.45)	292	13.47 (2.59)
Income-to-needs ratio	1,063	3.72 (2.82)	239	3.17 (3.03)
	N	*n* (%)	*N*	*n* (%)
Child sex (male)	1,071	536 (50)	293	169 (58)
Child ethnicity (European American non-Hispanic)	1,071	833 (78)	293	209 (71)
Partner in home (yes)	1,068	807 (76)	274	185 (68)
Study measures (time point, source)	*N*	*M* (*SD*)	*N*	*M* (*SD*)
Attachment security (36 months, observed)	983	5.07 (1.70)	157	4.78 (1.83)
Affective mutuality (54 months, observed)	965	5.22 (1.25)	75	4.77 (1.54)
Hostile attribution bias (54 months, child)	938	1.70 (1.33)	75	1.96 (1.31)
Hostile attribution bias (Grade 1, child)	974	4.70 (2.05)	51	4.53 (1.77)
Language ability (54 months, child)	979	101.28 (14.61)	81	98.07 (15.89)
Language ability (Grade 1, child)	970	106.36 (16.09)	50	105.16 (15.94)
Peer competence (Grade 1, mother)	980	15.67 (2.70)	49	16.08 (2.67)
Peer competence (Grade 1, teacher)	956	15.31 (3.63)	45	15.13 (3.22)
Positive friendship (Grade 3, mother)	1,008	3.22 (0.34)	—	
Negative friendship (Grade 3, mother)	1,008	1.97 (0.48)	—	
Positive friendship (Grade 3, teacher)	955	3.15 (0.39)	—	
Negative friendship (Grade 3, teacher)	960	1.89 (0.53)	—	

Note. Cases were included in this report if complete data were available for mother- and/or teacher-reported friendship quality in Grade 3 (*N* = 1,071). For this reason, no excluded cases had data on the friendship outcomes. Data on maternal age and education were collected at 1 month. Data on income-to-needs ratio were collected at 6, 15, 24, and 36 months and combined across time points. Data on the presence of a partner in the home were collected at 1, 6, 15, 24, and 36 months and combined across time points.

groups that parallel the infant classificatory system. In addition to classifying children as secure, avoidant, dependent/resistant, and controlling/disorganized, coders rated global security on a 9-point scale, ranging from 1 (*highly insecure*; e.g., highly avoidant, ambivalent, or disorganized or displaying a combination of strategies) to 9 (*highly secure*; e.g., highly at ease in initiating interaction, proximity, or contact; during reunion, child is calm but pleased to see mother return). Three coders were trained and certified by Jude Cassidy. Prior to official coding, all coders assessed the ABCD classifications on a set of 21 test tapes and achieved at least 75% agreement with Dr. Cassidy. For official coding, interobserver reliability was assessed on 75% of the protocols, and disagreements were resolved by consensus. For the attachment classifications (ABCD), intercoder agreement (before consensus) was 75.7% ($\kappa = .58$). For the global rating of security, intercoder agreement assessed via intraclass correlations (Winer, 1971) was .73.

An assessment of construct validity revealed consistent, albeit modest, associations with multiple indicators of maternal and child functioning (NICHD ECCRN, 2001). Moreover, post hoc comparisons among the four attachment groups revealed differences between the secure group and the various insecure groups; no differences emerged among the insecure groups. Because this measure shows validity in distinguishing between secure and insecure attachment, only the continuous rating of security was used in this report. We examined the security continuum (vs. binary score) because it provides a more fine-grained assessment of security.

Affective mutuality. During a laboratory visit at 54 months, mother–child interaction was videotaped during a 15-min semistructured observation. Mothers and children were presented with three activities: (a) completing a maze using an Etch A Sketch, (b) building a series of same-sized rectangular cube towers from variously shaped blocks, and (c) playing with a set of six hand puppets. The first two activities aimed to capture mother–child interaction during challenging situations in which the tasks were too difficult for the child to complete on his or her own, and the third activity aimed to capture mother–child play.

Videotapes of the mother–child interactions were coded for mother and child behaviors (see NICHD

ECCRN, 2003). Given the emphasis on attachment security in this report, we examined the rating of dyadic affective mutuality, which was based on work by Pianta (1994). Affective mutuality was coded on a 7-point scale, ranging from 1 (*very low*) to 7 (*very high*), and captured the degree to which mothers and children engaged in open verbal and nonverbal exchange of emotions. High affective mutuality characterized dyads in which the mother acknowledged the child's emotions and vice versa, the mother and child shared positive and negative experiences, and the child was free to express positive or negative feelings. In contrast, low affective mutuality characterized dyads in which the mother dampened the child's expression of emotion, and mother – child exchanges, especially around negative emotions, were nonreciprocal or stifled. Approximately 23% of the tapes were randomly assigned to two coders. Interobserver reliability, assessed via intraclass correlations (Winer, 1971), was .87. Mother – child affective mutuality as measured here has been related to more peer social skills and fewer behavior problems as reported by kindergarten teachers (Pianta et al., 1997).

Hostile Attribution Biases

Fifty-four months. Children's hostile attribution biases were assessed during a 54-month laboratory visit. Children were presented with four stories (without pictures) involving a same-sex peer: (a) having a special toy taken by another child, (b) being hit by a ball, (c) being tripped, and (d) having grape juice spilled on the child. For each story, the intent of the peer was ambiguous, and the child was asked to choose from one of two attributions (e.g., "Did Tom hit you in the back by accident or did Tom want to hit you in the back?"). Children received a score of 1 for indicating that the peer's action was intended, and responses were summed across vignettes to form a *hostile attribution* score (α = .64). This measure, adapted from the cartoon version developed by Dodge et al. (1986), was modified specifically for use with 4-year-old children (Feshbach, 1990). Hostile attributions assessed via this method have been associated with behavior problems, social competence, and child – friend interaction (Feshbach, 1990; Webster-Stratton & Lindsay, 1999).

First grade. During a laboratory visit at first grade, the Attribution Bias Interview was administered (Dodge et al., 1986). Children were presented with eight cartoon drawings of peer provocations, in which the intent of the peer was ambiguous. For each situation, children were asked how they would interpret the peer's intent ("Why do you think this happened?") and how they would respond in such a situation ("What would you do?"). The first question assessed attributions of intent and was coded as don't know, nonhostile, or hostile. A hostile attribution score was computed by summing hostile responses across stories (α = .64). This measure of hostile attributions has been associated with greater aggressive behavior and peer rejection (e.g., Dodge et al., 1986) and lower levels of prosocial behavior (Nelson & Crick, 1999).

Language Ability

The Woodcock – Johnson Psychoeducational Battery – Revised (WJ – R) was administered to children during laboratory visits at 54 months and first grade. The WJ – R consists of two batteries: Tests of Cognitive Abilities (WJ – R COG) and Tests of Achievement. The Picture Vocabulary Test (Test 6 of WJ – R COG) was examined for purposes of the current report. The first six items were presented in a multiple-choice format, in which children were asked to point to a named object. For subsequent items, children were asked to name pictured objects. The WJ – R has been normed on a nationally representative sample of 6,359 participants ranging in age from 24 months to 95 years and has shown good internal consistency and test – retest reliability (McGrew, Werder, & Woodcock, 1991). Standardized scores were examined in the analyses.

Peer Competence

The Social Skills Questionnaire from the Social Skills Rating System (SSRS; Gresham & Elliott, 1990) was administered to mothers (Parent Form) and teachers (Teacher Form) at Grade 1. This questionnaire consists of 49 items assessing the frequency of social behaviors, and items were rated on a 3-point scale, ranging from 0 (*never*) to 2 (*very often*). A peer competence subscale was computed by summing 10 items that tapped children's prosocial behavior toward peers and appropriate response to negative peer interactions (αs = .74 and .85, mothers and teachers, respectively). The SSRS has been normed on large samples and shows good construct validity and reliability (Gresham & Elliott, 1990).

Friendship Quality

The Quality of Child's Friendship questionnaire was administered to mothers and teachers at third grade. This 20-item questionnaire, which assesses the child's relationship with a best friend, was adapted from the Quality of Classroom Friends questionnaire

(Clark & Ladd, 2000). Twelve items assessed positive interaction ($\alpha = .88$ for both mothers and teachers) and 8 items assessed negative interaction (αs = .85 and .86, mothers and teachers, respectively). Items were rated on a 4-point scale, ranging from 1 (*strongly disagree*) to 4 (*strongly agree*). Subscales were computed by averaging across items. Clark and Ladd (2000) reported that kindergarten children who had more connected parent – child relationships tended to have more harmonious relationships with their closest friend as rated by teachers.

Data Analytic Strategy

Mplus 5.0 (Muthén & Muthén, 1998 – 2007) was used to test the hypothesized models. Full information maximum likelihood (FIML) was used to handle missing data. FIML offers less biased estimates compared with other methods such as listwise deletion (Enders & Bandalos, 2001; Schafer & Graham, 2002). To avoid inflating predictor – outcome associations, however, only cases with complete data on mother- and/or teacher-reported friendship quality were used in the analyses. The comparative fit index (CFI) and root mean square error of approximation (RMSEA) were examined to assess model fit. CFI values of 0.90 and above indicate an adequate fit, and values of 0.95 and above indicate a good fit (Hu & Bentler, 1995). RMSEA values less than 0.08 indicate adequate fit, and values less than 0.05 indicate a good fit (Browne & Cudeck, 1993). Maternal education and family income-to-needs ratio were included as covariates in all models presented below.

Model A and Model B depicted in Figure 1 were each estimated for mother and teacher reports of peer functioning (peer competence in Grade 1 and friendship quality in Grade 3). In testing Model B, three paths were added to Model A: 54-month hostile attributions, language ability, and mother – child affective mutuality to Grade 1 peer competence. Chi-square values for the two models were compared to determine the more plausible model. Error variances for manifest variables within the 54-month and Grade 1 time points were allowed to correlate, although for ease of presentation, the within-time covariances are not shown. For each model, the two subscales of the Quality of Child's Friendship questionnaire were indicators of a latent variable of friendship quality. Item ratings for the negative interaction subscale were reflected and averaged so that higher scores indicated greater friendship quality. For the more plausible model (Model A vs. Model B), total and specific indirect effects between child – mother attachment security at 36 months and friendship quality at Grade 3 were tested in Mplus using the delta method (i.e., Sobel, 1982). Because a secondary objective was to assess whether preschool attachment was associated with the proposed mechanisms via open emotional communication of the mother – child dyad, we also examined indirect effects between 36-month attachment security and Grade 1 mechanisms via mother – child affective mutuality at 54 months.

Two sets of follow-up analyses were conducted for the more plausible model. First, to examine child gender as a potential moderator of the indirect effects between attachment security and friendship quality, multigroup analyses were conducted. If a multigroup analysis indicated a significant gender effect, we assessed whether specific indirect effects between attachment security and friendship quality differed for boys and girls. Second, we examined mothers and teachers in the same model to compare path estimates predicting mother- versus teacher-reported peer competence in Grade 1 and friendship quality in Grade 3.

In assessing the proposed process model, it is important to distinguish between mediated and indirect effects. A mediated effect, which is considered a special case of indirect effects (Holmbeck, 1997; Preacher & Hayes, 2004), requires a significant association between the predictor and the outcome (Baron & Kenny, 1986). An indirect effect, in contrast, does not necessitate significance of this path. Rather, a significant indirect effect implies that the predictor has an effect on the outcome via the intervening variable (Kenny, Kashy, & Bolger, 1998). An examination of indirect effects may be especially important for developmental researchers examining processes that are distal in time (Shrout & Bolger, 2002). In this vein, we conducted a preliminary test of a "direct effects" model in which 36-month attachment security was the sole predictor of Grade 3 friendship quality (controlling for maternal education and family income). These analyses indicated that for both mother- and teacher-reported friendship quality, the direct effect of 36-month attachment security on friendship quality was nonsignificant. Thus, in the analyses that follow, we refer to indirect rather than mediated effects. Although Baron and Kenny's (1986) method for assessing mediation has been prevalent in the psychological literature, the importance of indirect effects, especially when examined across a long time span, should not be discounted because it is the predictor – mediator and mediator – outcome associations (vs. the predictor – outcome association) that are central to testing hypotheses concerning intervening mechanisms (Kenny et al., 1998).

Results

Preliminary Analyses

Cases included in the current report ($N = 1{,}071$) were compared to cases excluded due to missing data on either mother- or teacher-reported friendship quality in third grade (see Table 1 for descriptive statistics). Comparisons indicated that, on average, mothers included in the subsample were older, $t(1{,}362) = 4.74, p < .001$, and had received more years of education, $t(1{,}361) = 5.94, p \leq .001$. Participating families had a higher income-to-needs ratio, $t(1{,}300) = 2.67, p < .01$, and were more likely to have a partner living in the home, $\chi^2(1, N = 1{,}342) = 7.32, p < .01$. Furthermore, children included in the subsample were more likely to be female, $\chi^2(1, N = 1{,}364) = 5.37, p < .05$, and European American, $\chi^2(1, N = 1{,}364) = 5.30, p < .05$. Comparisons on the study measures indicated that included cases had higher levels of child–mother attachment security at 36 months, $t(1{,}138) = 1.96, p = .05$, and mother–child affective mutuality at 54 months, $t(1{,}038) = 2.94, p < .01$ (see Table 1 for means). Cases included versus excluded did not differ significantly on hostile attributions or language ability at 54 months and first grade nor mother- or teacher-reported peer competence at first grade.

For descriptive purposes, correlations among the study measures are presented in Table 2. Bivariate correlations are shown above the diagonal, and partial correlations controlling for maternal education and family income-to-needs ratio are shown below the diagonal.

Tests of the Process Model

The hypothesized process model depicted in Model A (see Figure 1) was tested separately for mother and teacher reports. Chi-square tests were significant, which would be expected given the large sample size: $\chi^2(20) = 68.57, p < .001$, and $\chi^2(20) = 54.65, p < .001$, mother and teacher models, respectively. Fit indices for Model A indicated good fit: RMSEAs = 0.048 and 0.040, CFIs = 0.973 and 0.978, mother and teacher models, respectively. Next, we tested Model B (see Figure 1) in which paths from 54-month hostile attributions, language ability, and mother–child affective mutuality to Grade 1 peer competence were added to Model A. Chi-square tests were again significant: $\chi^2(17) = 44.08, p < .001$, and $\chi^2(17) = 34.52, p < .01$, mother and teacher models, respectively. Fit indices for Model B indicated good fit for both mother and teacher models: RMSEAs = 0.039 and 0.031, CFIs = 0.985 and 0.989. Comparisons between Model A and Model B indicated that Model B showed significantly improved fit over Model A, $\chi^2_{\text{diff}}(3) = 24.49$ and 20.13, $ps < .001$, mother and teacher models, respectively. Because Model B was the more plausible model, we proceeded by examining the parameter estimates and testing indirect effects for Model B only. Standardized path estimates and R^2 values for mother and teacher models are presented in Figure 2. For paths not involving mother or teacher reports (e.g., path from 36-month attachment security to 54-month hostile attributions), path estimates were identical or nearly identical (i.e., differed at the third decimal place) across the mother and teacher models. Thus, for those paths, only one

Table 2

Intercorrelations Among Study Measures

Study measure: Time point, source	1	2	3	4	5	6	7	8	9	10	11	12
1. Attachment security: 36 months, observed	—	.24***	−.06	.17***	−.09**	.18***	.12***	.14***	.06	−.02	.06	−.08*
2. Affective mutuality: 54 months, observed	.21***	—	−.10**	.22***	−.14***	.26***	.17***	.13***	.15***	−.17***	.13***	−.12***
3. Hostile attribution bias: 54 months, child	−.04	−.07*	—	−.05	.12***	−.08*	−.02	−.07*	−.09**	.05	−.05	.05
4. Language ability: 54 months, child	.11***	.13***	.00	—	−.17***	.67***	.19***	.17***	.19***	−.15***	.11***	−.13***
5. Hostile attribution bias: Grade 1, child	−.08*	−.12***	.11**	−.13***	—	−.15***	−.12***	−.13***	−.10**	.09**	−.13***	.14***
6. Language ability: Grade 1, child	.13***	.17***	−.04	.59***	−.11***	—	.22***	.18***	.17***	−.17***	.11**	−.14***
7. Peer competence: Grade 1, mother	.10**	.13***	.01	.10**	−.10**	.14***	—	.29***	.35***	−.26***	.18***	−.15***
8. Peer competence: Grade 1, teacher	.12***	.10**	−.05	.11***	−.12***	.12***	.26***	—	.24***	−.21***	.29***	−.26***
9. Positive friendship: Grade 3, mother	.02	.09**	−.06	.09**	−.08*	.08*	.31***	.21***	—	−.66***	.22***	−.20***
10. Negative friendship: Grade 3, mother	.01	−.13***	.03	−.08*	.07*	−.10**	−.22***	−.18***	−.64***	—	−.20***	.20***
11. Positive friendship: Grade 3, teacher	.03	.10**	−.04	.04	−.11***	.04	.15***	.27***	.19***	−.17***	—	−.59***
12. Negative friendship: Grade 3, teacher	−.06	−.08*	.04	−.07*	.12***	−.08*	−.12***	−.25***	−.17***	.18***	−.58***	—

Note. Bivariate correlations are reported above the diagonal and partial correlations controlling for maternal education and family income-to-needs ratio are reported below the diagonal.

$^*p < .05$. $^{**}p < .01$. $^{***}p < .001$.

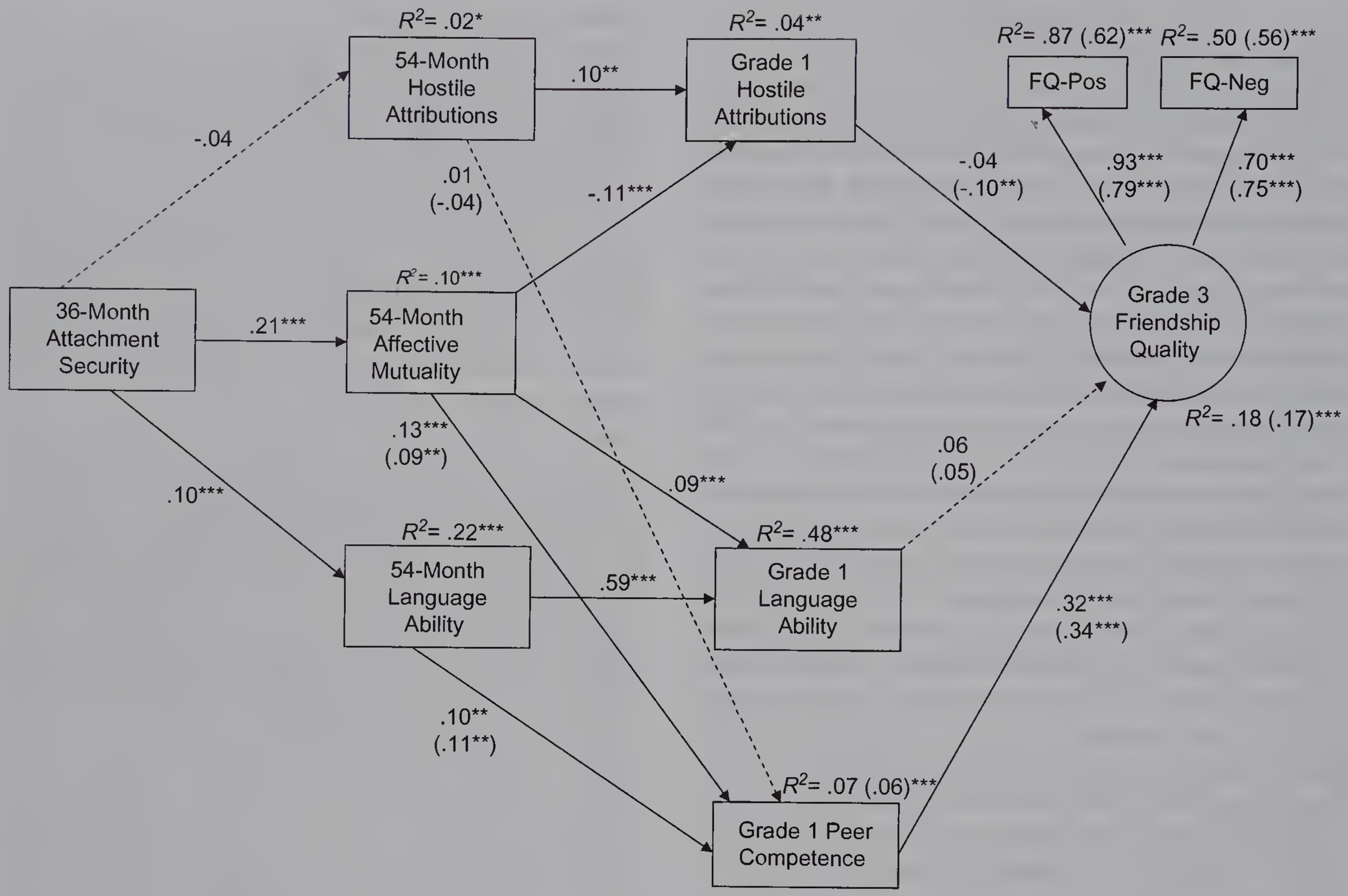

R^2= .02*
54-Month Hostile Attributions
.10**
R^2= .04**
Grade 1 Hostile Attributions
R^2= .87 (.62)***
R^2= .50 (.56)***
FQ-Pos
FQ-Neg
-.04
.01 (-.04)
-.11***
-.04 (-.10**)
.93*** (.79***)
.70*** (.75***)
R^2= .10***
36-Month Attachment Security
.21***
54-Month Affective Mutuality
Grade 3 Friendship Quality
R^2= .18 (.17)***
.13*** (.09**)
.06 (.05)
.09***
.10***
R^2= .22***
54-Month Language Ability
.59***
R^2= .48***
Grade 1 Language Ability
.32*** (.34***)
.10** (.11**)
R^2= .07 (.06)***
Grade 1 Peer Competence

Figure 2. Hostile attributions, language ability, affective mutuality, and peer competence as intervening mechanisms between child – mother attachment at 36 months and friendship quality at Grade 3.

Note. Models (N = 1,071) were estimated separately for mother and teacher reports of peer functioning using full information maximum likelihood. Item ratings for FQ-Neg were reflected and averaged so that higher scores indicated greater friendship quality. For those paths predicting peer competence in Grade 1 and friendship quality in Grade 3, standardized path estimates for mother reports are presented outside parentheses, and standardized path estimates for teacher reports are presented inside parentheses. For all other paths, estimates were identical across mother and teacher models, and only one estimate is shown. Paths represented by dotted lines were nonsignificant for both mother and teacher models. R^2 estimates are presented for each endogenous variable. Maternal education and family income-to-needs ratio were included as covariates in the models. For ease of presentation, error terms, within-time covariances, and covariates are not shown. A table including these additional parameter estimates is available from the first author. FQ-Pos = friendship quality, positive interaction subscale; FQ-Neg = friendship quality, negative interaction subscale.
$^*p < .05$. $^{**}p < .01$. $^{***}p < .001$.

estimate (applicable to both mother and teacher models) is presented in Figure 2. For paths to mother- and teacher-reported peer competence in Grade 1 and friendship quality in Grade 3, estimates for the mother model are presented outside parentheses, and estimates for the teacher model are presented inside parentheses (see Figure 2).

The test of the total indirect effect (i.e., all specific indirect effects combined) from 36-month attachment security to Grade 3 friendship quality was significant for mother (z = 3.72, $p < .001$) and teacher (z = 3.40, p = .001) models. Next, we summarize results for each intrapersonal mechanism, in turn, with a focus on the tests of specific indirect effects between (a) attachment security and Grade 3 friendship quality and (b) attachment security and Grade 1 mechanisms via mother – child affective mutuality at 54 months.

Hostile attributions. As seen in Figure 2, child – mother attachment security at 36 months did not predict 54-month hostile attributions, although attachment security was indirectly related to Grade 1 hostile attributions via mother – child affective mutuality at 54 months (zs = −2.87 and −2.88, ps < .01, mother and teacher models, respectively). Furthermore, Grade 1 hostile attributions predicted lower friendship quality as reported by teachers in Grade 3, and this specific indirect effect (36-month attachment security→54-month affective mutuality→Grade 1 hostile attributions→Grade 3 teacher-reported friendship quality) was significant (z = 1.98,

$p < .05$). In contrast to the teacher model, Grade 1 hostile attributions were not significantly related to mother-reported friendship quality in Grade 3.

Language ability. Greater child–mother attachment security at 36 months was associated with greater language ability at 54 months, which, in turn, was associated with greater language ability at Grade 1 (see Figure 2). Moreover, 36-month attachment security had an indirect effect on Grade 1 language ability via mother–child affective mutuality at 54 months ($zs = 3.24$ and 3.26, $ps = .001$, mother and teacher models, respectively). Language ability in Grade 1, however, was not associated with friendship quality in Grade 3 as reported by mothers or teachers. Therefore, specific indirect effects of attachment security on friendship quality via Grade 1 language ability were nonsignificant.

Peer competence. For both mother and teacher models, affective mutuality and language ability, but not hostile attributions, at 54 months were associated with Grade 1 peer competence (see Figure 2). Furthermore, indirect effects of attachment security on peer competence emerged via 54-month affective mutuality ($zs = 3.26$ and 2.48, $ps < .001$ and .05, mother- and teacher-reported peer competence, respectively) and 54-month language ability ($zs = 2.14$ and 2.26, $ps < .05$). Grade 1 peer competence, in turn, was associated with greater mother- and teacher-reported friendship quality in Grade 3 (see Figure 2). Tests of the following indirect effects were significant: (a) 36-month attachment security → 54-month affective mutuality → Grade 1 peer competence → Grade 3 friendship quality ($zs = 3.09$ and 2.37, $ps < .01$ and .05, mother and teacher models, respectively) and (b) 36-month attachment security → 54-month language ability → Grade 1 peer competence → Grade 3 friendship quality ($zs = 2.09$ and 2.17, $ps < .05$).

Follow-Up Analyses

Child gender. We examined gender as a moderator of Model B. Multigroup analyses indicated that the unconstrained versus constrained model provided an improved fit for the mother model only, $\chi^2_{\text{diff}}(27) = 45.56$, $p < .05$. Given the effect of child gender for the mother model, we proceeded to examine specific indirect pathways for boys and girls. Comparisons revealed that the following indirect effects were significant for boys but not for girls: (a) attachment security → affective mutuality → peer competence, $\chi^2_{\text{diff}}(2) = 8.68$, $p < .05$ ($z = 3.34$, $p = .001$, for boys; $z = 0.85$, *ns*, for girls) and (b) attachment security → affective mutuality → peer competence → friendship quality, $\chi^2_{\text{diff}}(3) = 14.47$, $p < .01$ ($z = 2.73$, $p < .01$, for boys; $z = 0.85$, *ns*, for girls). Inspection of the individual paths composing these indirect effects indicated that the path from mother–child affective mutuality to mother-reported peer competence was significant for boys only ($\beta = .23$, $p < .001$, for boys; $\beta = .04$, *ns*, for girls).

Mother–teacher comparisons. Next, mother and teacher reports of peer competence and friendship quality were examined in one model to assess whether the strength of the path estimates differed by reporter. For a given predictor–outcome association (e.g., Grade 1 hostile attributions to Grade 3 friendship quality), mother and teacher paths were constrained to be equal, and we assessed whether model fit showed a significant improvement when the paths were free to vary. Of the six comparisons made (three paths to mother- vs. teacher-reported peer competence in Grade 1 and three paths to mother- vs. teacher-reported friendship quality in Grade 3), no significant differences emerged.

Discussion

Early attachment relationships are hypothesized to guide children's later interactions with others, especially close others, via internal working models of the self, other, and relationships. Despite the centrality of this premise to attachment theory, little is known about mechanisms that may account for associations between child–mother attachment security and children's subsequent close relationships with peers. To this end, we tested models of attachment–friend linkages using data from the NICHD SECCYD. Our hypothesized model, guided by attachment theory, posited hostile attribution biases and language ability as intrapersonal mechanisms through which attachment security at 36 months would be associated with children's friendship quality in third grade (see Model A, Figure 1). An alternative model considered peer competence in Grade 1 as part of the chain of intervening mechanisms (see Model B, Figure 1). Models were tested separately for mother- and teacher-reported peer functioning (peer competence and friendship quality), and in each case, the second model (Model B) provided a better fit to the data. Because open emotional communication is viewed as integral to attachment security, a secondary aim was to examine mother–child affective mutuality at 54 months as an intervening interpersonal mechanism between 36-month attachment security and Grade 1 intrapersonal mechanisms.

In accord with the notion of the internal working model as an interpretative filter (e.g., Main et al., 1985), we investigated children's hostile attribution biases as a potential mechanism of attachment – friend linkages. This hypothesis received partial support. Namely, children who experienced greater attachment security with mothers at 36 months tended to engage in more open emotional communication with mothers at 54 months. Open emotional communication at 54 months, in turn, was associated with fewer hostile attributions at Grade 1, which predicted greater teacher-reported friendship quality at Grade 3. Notably, in addition to controlling for maternal education and family income, the association between 54-month affective mutuality and Grade 1 hostile attributions controlled for 54-month hostile attributions, and the association between Grade 1 hostile attributions and Grade 3 friendship quality controlled for Grade 1 language ability and peer competence. As such, our longitudinal, repeated measures strategy provided a stringent test of hostile attributions as an intervening mechanism and lends confidence to the direction of effects posited by the process model. Moreover, these findings are consistent with prior evidence that child – mother attachment security is associated with fewer hostile attributions (Cassidy et al., 1996; Suess et al., 1992) and that hostile attributions are associated with problematic peer functioning (e.g., Dodge et al., 2003; Feldman & Dodge, 1987; Webster-Stratton & Lindsay, 1999). To our knowledge, the current study is the first to examine hostile attributions as a mechanism linking early child – mother attachment with later friendship quality. For the mother model, this indirect effect was nonsignificant, and specifically, Grade 1 hostile attributions were not associated with mother-reported friendship quality, above and beyond language, peer competence, and family demographics. Because vignettes assessing hostile attributions in first grade depicted peer provocations in a school context, Grade 1 hostile attributions may be a more robust predictor of teacher-reported friendship quality. That being said, however, a mother – teacher comparison of this path was nonsignificant.

Surprisingly, attachment security was not associated with 54-month hostile attributions. Likewise, 54-month hostile attributions showed a weak association with Grade 1 hostile attributions and no association with mother- or teacher-reported peer competence. Hostile attributions as assessed at 54 months have been associated in expected ways with teacher-reported social competence and observed child – friend interaction (Webster-Stratton & Lindsay, 1999), yet this prior work assessed bivariate associations only, and participants were predominantly male, at higher risk, and slightly older (68 months, on average). This measure of hostile attribution biases may be less suitable for younger, lower risk children. Also recall that the assessments of hostile attributions varied slightly at the two time points: At 54 months, no picture cues were provided and a forced-choice response format was used, whereas at first grade, line drawings accompanied the vignettes and an open-ended response format was used. Interestingly, in this respect, Cassidy et al. (1996, see study 2) reported attachment-related differences when hostile attributions were assessed via an open-ended question ("Why did the boy/girl . . . ?"), but not a forced-choice question ("Do you think that the boy/girl . . . on purpose or . . . by accident?"). Perhaps children are more likely to rely on their preexisting knowledge base (i.e., latent mental structures) when asked an open-ended question about an actor's intent. A forced-choice question, on the other hand, provides children with socially acceptable or even novel response options and, as such, may mask detection of meaningful individual differences.

A second intervening mechanism of interest was child language ability. For mother and teacher models, child – mother attachment security at 36 months predicted greater language ability at 54 months, which predicted greater peer competence at Grade 1. Peer competence, in turn, predicted greater friendship quality at Grade 3. This indirect effect provides support for language as an intervening mechanism of attachment – friend linkages, but one in which early language ability impacts later friendship quality via more general peer competence. That attachment security predicted language ability above and beyond family demographic variables adds to growing evidence of attachment as a communication system (Bowlby, 1988; Bretherton, 1990; Main et al., 1985) and context for language development (Belsky & Fearon, 2002; van IJzendoorn et al., 1995). The current results also converge with prior work suggesting the importance of language and communicative abilities for children's social skills and friendship quality (e.g., Farmer, 2006; Gottman, 1983). In short, poor language skills during the preschool years may set children on trajectories of problematic relationships with peers. Notably, Grade 1 language ability did not predict mother- or teacher-reported friendship quality above and beyond the other predictors in the model. It may be that language ability as assessed here was not a highly sensitive indicator of the types of language skills that are important for interactions with friends in middle childhood. Researchers should continue to elucidate the role of language as a mechanism of attachment – friend linkages, and adopting more fine-grained assessments of language skills may result in even stronger evidence of language as a critical process linking children's family and peer relationships.

A secondary aim of this study was to assess whether 36-month attachment security was associated indirectly with Grade 1 mechanisms via mother–child affective mutuality at 54 months. Results provided strong support for such processes. Namely, controlling for maternal education and family income, greater child–mother attachment security was associated with more open emotional exchange between 54-month-old children and their mothers. Open emotional exchange, in turn, was associated with fewer hostile attributions in Grade 1 (controlling also for hostile attributions at 54 months), greater language ability in Grade 1 (controlling also for language ability at 54 months), and greater peer competence in Grade 1 (controlling also for hostile attributions and language ability at 54 months). Thus, open emotional communication may be a key interpersonal process through which early attachment security is associated with the development of social-cognitive, linguistic, and social-emotional competencies. We speculate that the "closed" communication style of insecure dyads (especially with regard to negative emotions and events) may result in a hostile attribution bias because the parent explicitly communicates hostile attributions to the child or because the parent does *not* communicate with the child, and the child is left to make sense of the negative emotions or events on his or her own. Similarly, low levels of affective mutuality of insecure dyads may result in fewer opportunities for children to develop language and socioemotional competencies that will be important for later positive interactions with peers.

Examination of child gender as a moderator of the process model indicated that, on the whole, intervening mechanisms operated similarly for boys and girls. For the mother model, however, the indirect effects from attachment security to Grade 1 peer competence and Grade 3 friendship quality via 54-month affective mutuality were significant for boys but not for girls. Further probing indicated that the path from mother–child affective mutuality at 54 months to peer competence at Grade 1 was significant for boys only. Boys, compared with girls, are often socialized to suppress expression of negative emotions (Brody, 1999). Open, mutual exchange of emotions with caregivers, therefore, may be particularly salient for young boys' learning of social-emotional skills that foster harmonious, effective peer interactions.

Strengths of the current research include the large sample size, independent assessments of constructs, and inclusion of maternal education and family income as covariates. These factors yield increased power and reliability to detect indirect effects (see Frazier, Tix, & Barron, 2004). Most importantly, perhaps, the longitudinal design reflected the theoretically guided temporal ordering of constructs in the process model, and in some instances, the repeated assessment of constructs ruled out the possibility that longitudinal associations were due to stability of constructs across time. Thus, greater confidence in conclusions about the proposed direction of effects was achieved by adopting a replicative strategy, which permits an examination of change over time (see Cole & Maxwell, 2003; Hoyle & Robinson, 2004).

Nonetheless, several caveats should be noted. The subsample examined here, although relatively diverse, differed from cases excluded (due to missing data on the Grade 3 friendship outcomes) on demographic and study measures. On the whole, children in the subsample tended to show higher functioning, and it is unclear whether the findings extend to more at-risk populations. Given the potential implications of this process model for intervention, inclusion of clinical populations, in particular, will be important. Also, the amount of variance explained was modest. Future studies should expand on the model presented here to examine additional mechanisms, including emotional regulatory processes that are likely to be shaped by attachment relationships (Cassidy, 1994), integrated into the processing of social information (Lemerise & Arsenio, 2000), and vital to peer functioning (Fabes & Eisenberg, 1992; Parker & Gottman, 1989). Moreover, because the quality of child–friend interaction is influenced by characteristics of both relationship partners (McElwain & Volling, 2002), it is likely that explained variance for friendship quality will increase when social-cognitive and language abilities of both children in the friendship dyad are considered. Finally, the proposed processes may emerge for some groups but not for others. In this report, child gender was examined as a moderator, although overall, the process model functioned similarly for boys and girls. In considering other moderators, investigation of developmental period may be fruitful. Associations between social-cognition and behavioral outcomes become stronger in later childhood and adolescence (e.g., Lansford et al., 2006), and the mechanisms examined here may become more pronounced later in development.

In sum, the current study extends prior work on attachment–friend associations by identifying chains of intervening mechanisms whereby child–mother attachment during the preschool period was associated with children's relationships with friends during middle childhood. Our findings suggest the utility of investigating such processes at the intersection of SIP, language, and mother–child communication about emotions.

References

Ainsworth, M. D. S., Blehar, M. C., Waters, E., & Wall, S. (1978). *Patterns of attachment: A psychological study of the Strange Situation*. Hillsdale, NJ: Erlbaum.

Baldwin, M. W. (1992). Relational schemas and the processing of social information. *Psychological Bulletin, 112*, 461–484.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173–1182.

Belsky, J., & Cassidy, J. (1994). Attachment: Theory and evidence. In M. Rutter & D. Hays (Eds.), *Development through life* (pp. 373–402). Oxford, England: Blackwell.

Belsky, J., & Fearon, R. (2002). Infant-mother attachment security, contextual risk, and early development: A moderational analysis. *Development and Psychopathology, 14*, 293–310.

Berndt, T. J. (2002). Friendship quality and social development. *Current Directions in Psychological Science, 11*, 7–10.

Bowlby, J. (1982). *Attachment and loss: Vol. 1. Attachment*. New York: Basic Books. (Original work published 1969)

Bowlby, J. (1988). *A secure base: Clinical applications of attachment theory*. London: Routledge.

Bretherton, I. (1990). Open communication and internal working models: Their role in the development of attachment relationships. In R. A. Thompson (Ed.), *Nebraska symposium on motivation: Vol. 36. Socioemotional development* (pp. 57–113). Lincoln: University of Nebraska Press.

Bretherton, I., & Munholland, K. A. (1999). Internal working models in attachment relationships. In J. Cassidy & P. R. Shaver (Eds.), *Handbook of attachment: Theory, research, and clinical implications* (pp. 89–111). New York: Guilford.

Brody, L. (1999). *Gender, emotion and the family*. Cambridge, MA: Harvard University Press.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

Cassidy, J. (1994). Emotion regulation: Influences of attachment relationships. In N. A. Fox (Ed.), The development of emotion regulation: Biological and behavioral considerations. *Monographs of the Society for Research in Child Development, 59*(2–3, Serial No. 240), 228–249.

Cassidy, J., Kirsh, S. J., Scolton, K. L., & Parke, R. D. (1996). Attachment and representations of peer relationships. *Developmental Psychology, 32*, 892–904.

Cassidy, J., & Marvin, R. S., with the MacArthur Working Group on Attachment. (1992). *Attachment organization in preschool children: Procedures and coding manual*. Unpublished manual, University of Virginia.

Clark, K. E., & Ladd, G. W. (2000). Connectedness and autonomy support in parent-child relationships: Links to children's socioemotional orientation and peer relationships. *Developmental Psychology, 36*, 485–498.

Cohen, N. J. (2001). *Language impairment and psychopathology in infants, children, and adolescents*. Thousand Oaks, CA: Sage.

Cole, D. A., & Maxwell, S. E. (2003). Testing mediational models with longitudinal data: Questions and tips in the use of structural equation modeling. *Journal of Abnormal Psychology, 112*, 558–577.

Crick, N. R., & Dodge, K. A. (1994). A reformulation of social information-processing mechanisms in children's social adjustment. *Psychological Bulletin, 115*, 74–101.

Crittenden, P. M. (1990). Internal representational models of attachment relationships. *Infant Mental Health Journal, 11*, 259–277.

Dodge, K. A., & Frame, C. L. (1982). Social cognitive biases and deficits in aggressive boys. *Child Development, 53*, 620–635.

Dodge, K. A., Lansford, J. E., Burks, V. S., Bates, J. E., Pettit, G. S., Fontaine, R., et al. (2003). Peer rejection and social information-processing factors in the development of aggressive behavior problems in children. *Child Development, 74*, 374–393.

Dodge, K. A., Pettit, G. S., McClaskey, C. L., & Brown, M. M. (1986). Social competence in children. *Monographs of the Society for Research in Child Development, 51*(2, Serial No. 213).

Dodge, K. A., & Schwartz, D. (1997). Social information processing mechanisms in aggressive behavior. In D. M. Stoff & J. Breiling (Eds.), *Handbook of antisocial behavior* (pp. 171–180). New York: Wiley.

Elicker, J., Englund, M., & Sroufe, L. A. (1992). Predicting peer competence and peer relationships in childhood from early parent-child relationships. In R. D. Parke & G. W. Ladd (Eds.), *Family-peer relationships: Modes of linkage* (pp. 77–106). Hillsdale, NJ: Erlbaum.

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimations for missing data in structural equation models. *Structural Equation Modeling, 8*, 430–457.

Etzion-Carasso, A., & Oppenheim, D. (2000). Open mother-preschooler communication. Relations with early secure attachment. *Attachment and Human Development, 2*, 347–370.

Fabes, R. A., & Eisenberg, N. (1992). Young children's coping with interpersonal anger. *Child Development, 63*, 116–128.

Farmer, M. (2006). Language and the development of social and emotional understanding. In J. Clegg &

J. Ginsborg (Eds.), *Language and social disadvantage: Theory into practice* (pp. 74–92). Hoboken, NJ: Wiley.

Farrar, M. J., Fasig, L. G., & Welch-Ross, M. K. (1997). Attachment and emotion in autobiographical memory development. *Journal of Experimental Child Psychology, 67,* 389–408.

Feldman, E., & Dodge, K. A. (1987). Social information processing and sociometric status: Sex, age, and situational effects. *Journal of Abnormal Child Psychology, 15,* 211–227.

Feshbach, L. (1990). Aggression-conduct problems, attention-deficits, hyperactivity, play, and social cognition in four-year-old boys (Doctoral dissertation, University of Washington, 1989). *Dissertation Abstracts International, 50,* 5878.

Fish, M., & Pinkerman, B. (2003). Language skills in low-SES rural Appalachian children: Normative development and individual differences, infancy to preschool. *Applied Developmental Psychology, 23,* 539–565.

Frazier, P. A., Tix, A. P., & Barron, K. E. (2004). Testing moderator and mediator effects in counseling psychology research. *Journal of Counseling Psychology, 51,* 115–134.

Gentzler, A. L., Contreras-Grau, J. M., Kerns, K. A., & Weimer, B. L. (2005). Parent-child emotional communication and children's coping in middle childhood. *Social Development, 14,* 591–612.

Gottman, J. M. (1983). How children become friends. *Monographs of the Society for Research in Child Development, 48*(3, Serial No. 201).

Gresham, F. M., & Elliott, S. N. (1990). *The social skills rating system.* Circle Pines, MN: American Guidance Service.

Hartup, W. W. (1996). The company they keep: Friendships and their developmental significance. *Child Development, 67,* 1–13.

Hebert-Myers, H., Guttentag, C. L., Swank, P. R., Smith, K. E., & Landry, S. H. (2006). The importance of language, social, and behavioral skills across early and later childhood as predictors of social competence with peers. *Applied Developmental Science, 10,* 174–187.

Holmbeck, G. N. (1997). Toward terminological, conceptual, and statistical clarity in the study of mediators and moderators: Examples from the child-clinical and pediatric psychology literatures. *Journal of Consulting and Clinical Psychology, 65,* 599–610.

Howes, C., Droege, K., & Matheson, C. C. (1994). Play and communicative processes within long- and short-term friendship dyads. *Journal of Social and Personal Relationships, 11,* 401–410.

Hoyle, R. H., & Robinson, J. C. (2004). Mediated and moderated effects in social psychological research: Measurement, design, and analysis issues. In C. Sansome, C. Morf, & A. T. Panter (Eds.), *The Sage handbook of methods in social psychology* (pp. 213–233). Thousand Oaks, CA: Sage.

Hu, L. T., & Bentler, P. M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 76–99). Thousand Oaks, CA: Sage.

Kenny, D., Kashy, D. A., & Bolger, N. (1998). Data analysis in social psychology. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (pp. 233–265). New York: McGraw-Hill.

Kobak, R., & Duemmler, S. (1994). Attachment and conversation: Toward a discourse analysis of adolescent and attachment security. In K. Bartholomew & D. Perlman (Eds.), *Advances in personal relationships: Vol. 5. Attachment processes in adulthood* (pp. 121–149). London: Jessica Kingsley.

Laible, D. (2004). Mother-child discourse in two contexts: Links with child temperament, attachment security, and socioemotional competence. *Developmental Psychology, 40,* 979–992.

Lansford, J. E., Malone, P. S., Dodge, K. A., Crozier, J. C., Pettit, G. S., & Bates, J. E. (2006). A 12-year prospective study of patterns of social information processing problems and externalizing behaviors. *Journal of Abnormal Child Psychology, 34,* 709–718.

Leaper, C. (1991). Influence and involvement in children's discourse: Age, gender, and partner effects. *Child Development, 62,* 797–811.

Lemerise, E. A., & Arsenio, W. F. (2000). An integrated model of emotion processes and cognition in social information processing. *Child Development, 71,* 107–118.

Lieberman, A. F. (1977). Preschooler's competence with a peer: Relations with attachment and peer experience. *Child Development, 48,* 1277–1287.

Lieberman, M., Doyle, A.-B., & Markiewicz, D. (1999). Developmental patterns in security of attachment to mother and father in late childhood and early adolescence: Associations with peer relations. *Child Development, 70,* 202–213.

Locke, J. L. (2001). First communion: The emergence of vocal relationships. *Social Development, 10,* 294–308.

Main, M. (1995). Recent studies in attachment. In S. Goldberg, R. Muir, & J. Kerr (Eds.), *Attachment theory: Social, developmental, and clinical perspectives* (pp. 407–474). Hillsdale, NJ: Analytic Press.

Main, M., Kaplan, N., & Cassidy, J. (1985). Security in infancy, childhood, and adulthood: A move to the level of representation. In I. Bretherton & E. Waters (Eds.), Growing points of attachment theory and research. *Monographs of the Society for Research in Child Development, 50*(1–2, Serial No. 209), 66–104.

McElwain, N. L., Cox, M. J., Burchinal, M. R., & Macfie, J. (2003). Differentiating among insecure mother-infant attachment classifications: A focus on child-friend interaction and exploration during solitary play at 36 months. *Attachment and Human Development, 5,* 136–164.

McElwain, N. L., & Volling, B. L. (2002). Relating individual control, social understanding, and gender to child-friend interaction: A relationships perspective. *Social Development, 11,* 362–385.

McElwain, N. L., & Volling, B. L. (2004). Attachment security and parental sensitivity during infancy: Associations with friendship quality and false belief understanding at age four. *Journal of Social and Personal Relationships*, *21*, 639–667.

McGrew, K. S., Werder, J. K., & Woodcock, R. W. (1991). *WJ–R technical manual*. Allen, TX: DLM.

Mendez, J. L., Fantuzzo, J., & Cicchetti, D. (2002). Profiles of social competence among low-income African American preschool children. *Child Development*, *73*, 1085–1100.

Muller, D., Judd, C. M., & Yzerbyt, V. Y. (2005). When moderation is mediated and mediation is moderated. *Journal of Personality and Social Psychology*, *89*, 852–863.

Muthén, L. K., & Muthén, B. O. (1998–2007). *Mplus user's guide* (5th ed.). Los Angeles: Author.

National Institute of Child Health and Human Development Early Child Care Research Network. (1997). The effects of infant child care on infant-mother attachment security: Results of the NICHD Study of Early Child Care. *Child Development*, *68*, 860–879.

National Institute of Child Health and Human Development Early Child Care Research Network. (1999). Child care and mother-child interaction in the first 3 years of life. *Developmental Psychology*, *35*, 1399–1413.

National Institute of Child Health and Human Development Early Child Care Research Network. (2001). Child care and family predictors of preschool attachment and stability from infancy. *Developmental Psychology*, *37*, 847–862.

National Institute of Child Health and Human Development Early Child Care Research Network. (2003). Early child care and mother-child interaction from 36 months through first grade. *Infant Behavior and Development*, *26*, 345–370.

Nelson, D. A., & Crick, N. R. (1999). Rose-colored glasses: Examining the social information-processing of prosocial young adolescents. *Journal of Early Adolescence*, *19*, 17–38.

Orobio de Castro, B., Veerman, J. W., Koops, W., Bosch, J. D., & Monshouwer, H. J. (2002). Hostile attribution of intent and aggressive behavior: A meta-analysis. *Child Development*, *73*, 916–934.

Parker, J. G., & Gottman, J. M. (1989). Social and emotional development in a relational context. In T. J. Berndt & G. W. Ladd (Eds.), *Peer relationships in child development* (pp. 95–131). New York: Wiley.

Pianta, R. C. (1994). *Rating scales for parent-child interaction in preschoolers*. Unpublished coding manual, University of Virginia.

Pianta, R. C., Nimetz, S. L., & Bennett, E. (1997). Mother-child relationships, teacher-child relationships, and school outcomes in preschool and kindergarten. *Early Childhood Research Quarterly*, *12*, 263–280.

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers*, *36*, 717–731.

Renk, K., & Phares, V. (2004). Cross-informant ratings of social competence in children and adolescents. *Clinical Psychology Review*, *24*, 239–254.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*, 147–177.

Schneider, B. H., Atkinson, L., & Tardiff, C. (2001). Child-parent attachment and children's peer relations: A quantitative review. *Developmental Psychology*, *37*, 86–100.

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods*, *7*, 422–445.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. In S. Leinhart (Ed.), *Sociological methodology* (pp. 290–312). San Francisco: Jossey-Bass.

Sroufe, L. A., & Fleeson, J. (1986). Attachment and the construction of relationships. In W. W. Hartup & Z. Rubin (Eds.), *Relationships and development* (pp. 51–71). Hillsdale, NJ: Erlbaum.

Suess, G., Grossmann, K., & Sroufe, L. (1992). Effects of infant attachment to mother and father on quality of adaptation in preschool: From dyadic to individual organisation of self. *International Journal of Behavioral Development*, *15*, 43–65.

Thompson, R. A. (1998). Early sociopersonality development. In W. Damon (Series Ed.) & N. Eisenberg (Vol. Ed.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (5th ed., pp. 25–104). New York: Wiley.

Thompson, R. A. (2000). The legacy of early attachments. *Child Development*, *71*, 145–152.

Tomasello, M. (1992). The social bases of language acquisition. *Social Development*, *1*, 67–87.

van IJzendoorn, M. H., Dijkstra, J., & Bus, A. G. (1995). Attachment, intelligence, and language: A meta-analysis. *Social Development*, *4*, 115–128.

Webster-Stratton, C., & Lindsay, D. W. (1999). Social competence and conduct problems in young children: Issues in assessment. *Journal of Clinical Child Psychology*, *28*, 25–43.

Winer, B. J. (1971). *Statistical principles in experimental design* (2nd ed.). New York: McGraw-Hill.

Youngblade, L. M., & Belsky, J. (1992). Parent-child antecedents of 5-year-olds' close friendships: A longitudinal analysis. *Developmental Psychology*, *28*, 700–713.

Ziv, Y., Oppenheim, D., & Sagi-Schwartz, A. (2004). Social information processing in middle childhood: Relations to infant-mother attachment. *Attachment and Human Development*, *6*, 327–348.

Child Development, November/December 2008, Volume 79, Number 6, Pages 1907–1927

Testing an Idealized Dynamic Cascade Model of the Development of Serious Violence in Adolescence

Kenneth A. Dodge
Duke University

Mark T. Greenberg
Pennsylvania State University

Patrick S. Malone
University of South Carolina

Conduct Problems Prevention Research Group

A dynamic cascade model of development of serious adolescent violence was proposed and tested through prospective inquiry with 754 children (50% male; 43% African American) from 27 schools at 4 geographic sites followed annually from kindergarten through Grade 11 (ages 5–18). Self, parent, teacher, peer, observer, and administrative reports provided data. Partial least squares analyses revealed a cascade of prediction and mediation: An early social context of disadvantage predicts harsh–inconsistent parenting, which predicts social and cognitive deficits, which predicts conduct problem behavior, which predicts elementary school social and academic failure, which predicts parental withdrawal from supervision and monitoring, which predicts deviant peer associations, which ultimately predicts adolescent violence. Findings suggest targets for in-depth inquiry and preventive intervention.

Consensus grows that children who are at high risk for later chronic violence can be identified early in life based on conduct problems (called the early-starter model; Moffitt, 1993; Patterson, Reid, & Dishion, 1992), but the processes are less clear through which minor conduct problems in early life get transformed into costly and serious violent behaviors in adolescence. Numerous variables (classified into domains such as context, parenting, and peer relations; Stouthamer-Loeber, Loeber, Wei, Farrington, & Wikström, 2002) are known to predict adolescent antisocial outcomes, but whether these variables increment the prediction beyond dispositional factors in early life is still under debate (reviewed by Dodge, Coie, & Lynam, 2006). A developmental model of conduct disorder (Conduct Problems Prevention Research Group [CPPRG], 1992) has identified some of the major predictor domains and posits that these domains catalyze early conduct problems into new, more serious, chronically violent behaviors in adolescence. Numerous questions remain about the processes through which early variables operate to produce chronic violence (Dodge & Pettit, 2003). It is not clear how predictor domains relate to each other, whether one domain mediates the impact of another domain, whether some domains account for redundant portions of outcome variance, and whether each domain provides unique increments in predicting violence in adolescence. The major alternate hypothesis (Gottfredson & Hirschi, 1990) is that some

---

Members of the Conduct Problems Prevention Research Group are, in alphabetical order, Karen L. Bierman, PhD, Department of Psychology, Pennsylvania State University; John D. Coie, PhD, Department of Psychology, Duke University; Kenneth A. Dodge, PhD, Center for Child and Family Policy, Duke University; Mark T. Greenberg, PhD, Department of Human Development and Family Studies, Pennsylvania State University; John E. Lochman, PhD, Department of Psychology, University of Alabama; Robert J. McMahon, PhD, Department of Psychology, University of Washington; and Ellen E. Pinderhughes, PhD, Department of Child Development, Tufts University. This work was supported by National Institute of Mental Health (NIMH) Grants R18 MH48043, R18 MH50951, R18 MH50952, and R18 MH50953. The Center for Substance Abuse Prevention and the National Institute on Drug Abuse also provided support for Fast Track through a memorandum of agreement with NIMH. This work was also supported by U.S. Department of Education Grant S184U30002 and NIMH Grants K05MH00797 and K05MH01027. Although these funding agencies approved the initial design of the study, they had no role in approving the analysis and interpretation of the data or in preparation and review of the manuscript. We are grateful for the close collaboration of the Durham Public Schools, the Metropolitan Nashville Public Schools, the Bellefonte Area Schools, the Tyrone Area Schools, the Mifflin County Schools, the Highline Public Schools, and the Seattle Public Schools. We appreciate the dedication of hundreds of staff members who implemented the project, collected the evaluation data, and assisted with data management. Requests for reprints should be addressed to Seattle Fast Track, 146 North Canal Street, Suite 111, Seattle, WA 98103. For additional information concerning Fast Track, see http://www.fasttrackproject.org.

Correspondence concerning this article should be addressed to Kenneth A. Dodge, Center for Child and Family Policy, Box 90545, Duke University, Durham, NC 27708. Electronic mail may be sent to dodge@duke.edu.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7906-0025

children are born with, or acquire early in life, a disposition for low self-control and become early starters destined for violent outcomes; all other "predictor" variables represent noncausal markers of the child's impact on the world that do not play any incremental role in that child's development. The goals of the current study were to articulate a dynamic cascade model of the development of chronic violence in adolescence that posits incremental roles of numerous life experience factors and to test key hypotheses from the model through a prospective study of children followed from kindergarten through late adolescence. We acknowledge that other plausible cascade models could be tested as well.

Predictors of Adolescent Antisocial Behavior

A large body of empirical evidence has established the predictive validity of at least seven domains of predictors of adolescent conduct problems: early adverse social context, early harsh and inconsistent parenting, lack of social-cognitive and cognitive readiness to begin school, early behavior problems, early school social and academic failure, lack of parental supervision and monitoring in adolescence, and adolescent deviant peer associations (CPPRG, 1992; Dodge, Coie, et al., 2006; Stouthamer-Loeber et al., 2002). Further, it appears that most of these same factors also predict severe interpersonal violence and homicide, the most severe form of conduct problems (Capaldi & Patterson, 1996; Loeber et al., 2005). Other domains are plausible as well, and some of the following domains might be subdivided into more domains, but we offer the following as a heuristic model.

Early adverse social context. Using multilevel analyses of families residing in different neighborhoods in Chicago, Sampson, Raudenbush, and Earls (1997) found large neighborhood variations in youth violent behaviors that are associated with neighborhood structural characteristics. Beyers, Bates, Pettit, and Dodge (2003) used census tract data, and Ingoldsby and Shaw (2002) used self-reports to determine that neighborhood-level scores for the proportions of families characterized by poverty represent a significant risk factor for later individual-level conduct problems. Thus, the neighborhood social context into which a child is born places that child at high (or low) risk for violent behavior many years into the future (Stouthamer-Loeber et al., 2002). None of these scholars proposed that neighborhood context operates directly on a child's behavior; however, the processes through which context exerts influence are not clear.

Adverse context is also indexed by within-family variables during the first 5 years of life. Family socioeconomic status, which is fairly stable across childhood, exerts broad influence across domains from parenting to later conduct problems (McLoyd, 1990). Maternal depression affects the interactions between the mother and child, which may indirectly affect the trajectory of the child's behavioral development. Numerous other community and family context variables are extremely important but are not reviewed here because the focus of this study is on development during the school years.

It must be noted that the paths between adverse contexts and parenting are likely to be reciprocal. Parents with depression and other problems may suffer outcomes that lead them to migrate toward contexts of greater disadvantage.

Early harsh and inconsistent parenting. Most developmental theories propose a major role for the impact of early parenting strategies, especially during discipline encounters with a child (Patterson, 1986; Snyder, Reid, & Patterson, 2003). Harsh discipline in the first 5 years of life has been associated with violent behavior outcomes in adolescence (Lansford et al., 2002), especially when applied inconsistently (Patterson et al., 1992). The lack of warmth between parent and child is also crucial as a predictor (Pettit, Bates, & Dodge, 1997). Caspi et al. (2004) used a monozygotic twin study that controls for genetic differences and found that maternal verbal statements of negative affect about a child (called expressed emotion) predicted children's antisocial behavior problems.

McLoyd (1990) examined the parenting mechanisms through which a context of poverty might exert its impact. She found that poverty increases single African American parents' psychological distress and impairs their social support systems, which, in turn, diminishes their effectiveness and increases their coerciveness toward their children. Sampson and Laub (1994) reanalyzed the Glueck and Glueck (1950) longitudinal data set involving 1,000 Depression-era White families and found that the structural variable of family poverty influenced family processes of harsh discipline, low supervision, and poor parent–child attachment, which, in turn, influenced juvenile delinquency. Family process accounted for two thirds of the effect of poverty on delinquency. Likewise, Dodge, Pettit, and Bates (1994) found that harsh physical discipline by parents accounted for about half of the effect of low socioeconomic status on children's aggressive behavior. Numerous scholars have examined parenting processes during the first 5 years of life in more transactional detail, but the constraints of the design of the current study, which started when children were 5 years old, prohibited detailed analyses of this era.

Lack of school social and cognitive readiness. Children begin formal schooling with large individual differences in social-cognitive and cognitive skills and biases that have been shown to predict antisocial outcomes. Children who are unable to recognize emotions in others (Eisenberg et al., 2004), who demonstrate a bias to attribute hostile intent to peers' intentions (Dodge et al., 1986), and who are relatively poor at social problem solving (Dodge et al., 1986) are at risk for later antisocial behavior. Verbal abilities assessed by intelligence tests have also been found to predict later antisocial behavior (Moffitt & Caspi, 2001).

A growing literature has identified antecedents of social-cognitive skills, particularly in previous parenting. Pollak and Tolley-Schell (2003) found that physically abused children demonstrate selective attention to angry faces and reduced attention to happy faces. Dodge, Bates, and Pettit (1990) identified a pattern of hostile attributional bias and social problem-solving skill deficits that emerged following the experience of harsh discipline from parents. Aber, Gershoff, Ware, and Kotler (2004) found that witnessing family violence around the time of the September 11 disaster predicted children's tendencies to display hostile attributional biases a year later, and Cummings and Davies (2002) have argued that emotion dysregulation is the process that mediates the relation between witnessing marital conflict and later child aggression.

Early behavior problems. The continuity of externalizing behavior problems across development is one of the most well-documented findings in longitudinal inquiry in social development (Dodge, Coie, et al., 2006). Although the form of antisocial behavior may change from disobedience and oppositional behavior in early childhood to violence and crime in adolescence, continuity in the rank ordering of individual differences is the basis for the early-starter model (Moffitt, 1993). It is imperative that a cascade model describes the processes that transform antisocial behavior from one form to another and highlights factors that deflect this trajectory.

School social and academic failure. Deficits in school readiness skills are hypothesized to lead to school failure in peer relations and academic performance, which, in turn, have been linked to later antisocial outcomes. Dodge et al. (2003) found that the experience of social rejection by peers leads to growth in antisocial behavior across elementary school. Academic failure, indexed by grade retention, placement into special education, and failing test scores, represents another predictor of antisocial outcomes (Hinshaw, 1992; Moffitt, Gabrielli, Mednick, & Schulsinger, 1981). Meta-analyses suggest that kindergarten and early grade retention has long-term detrimental effects on behavior outcomes, in spite of temporary academic benefits (Holmes, 1989).

Lack of parental monitoring. As children move into adolescence, the tasks of parenting change, as do the parenting predictors of antisocial outcomes. Parents' lack of monitoring of their child's whereabouts, failure to supervise behavior and time allocation, and lack of engagement with the youth in schooling all have been associated with antisocial outcomes in adolescence (Dishion & McMahon, 1998). Monitoring is especially important as a protective factor among youth who reside in a social context of danger and risk (Pettit, Bates, Dodge, & Meece, 1999). Dishion, Nelson, and Bullock (2004) used video records of parent–adolescent interaction to determine that directly observed parental disengagement predicted growth in the youth's associations with deviant peers and delinquent outcomes. Hill et al. (2004) found that lack of parental involvement in school during the middle school years predicted a youth's later maladjustment.

Although individual differences in parents' lack of monitoring and engagement have been found to increment the prediction of antisocial outcomes beyond the levels that anteceded monitoring, it is plausible that these patterns are at least partly elicited by the child. Parent–child conflict may well be stressful, and monitoring by parents might intensify conflict when the parent discovers misbehavior. Thus, early school problems might well lead to lower levels of parental monitoring in adolescence (Stattin & Kerr, 2000). Laird, Pettit, Bates, and Dodge (2003) found evidence to support a reciprocal-influence model of growth in youth antisocial behavior and parental withdrawal of monitoring over time. Evidence by Larzelere and Patterson (1990) indicates that parent management practices in sixth grade, including monitoring, mediated the impact of prior family socioeconomic status on antisocial behavior measured in seventh grade from police records and self-report. Thus, parents' monitoring activities may well grow out of prior contexts but appear to increment and mediate later outcomes.

Deviant peer associations. The final domain of predictors of adolescent violent behavior is association with deviant peers. This factor has received a wide variety of empirical support as a proximal predictor of adolescent deviant behavior. Patterson et al. (1992) found that involvement with deviant peers in Grade 6 (measured by parent, peer, and self descriptions) predicted subsequent delinquency even controlling for prior antisocial behavior. Simons, Wu, Conger, and Lorenz (1994) also found that association with deviant peers predicted subsequent arrests. Laird, Jordan, Dodge, Pettit, and Bates (2001) found that deviant peer associations predicted growth in antisocial behavior, and Keenan, Loeber, Zhang, Stouthamer-Loeber, and Van Kammen (1995) found that in the Pittsburgh Youth Study authority conflict was twice as likely to occur among those disruptive boys who had truant or disobedient best friends as among those who did not. Two important features of this study were the fact that onset was the dependent variable, thus controlling for previous disruptive behavior, and the fact that peer influence was tested within a specific type of antisocial activity as a way of examining the extent to which peers truly were models of deviant behavior.

The antecedents of deviant peer associations are varied, but low parental monitoring has been implicated in several studies. Snyder, Dishion, and Patterson (1986) found a strong path relation between low parental monitoring and increases in deviant peer associations. Simons et al. (1994) found that problems in parental discipline practices (including poor monitoring) predicted increased deviant peer associations and later criminal arrests. Thus, it is plausible that the effect of the lack of parental monitoring on antisocial outcomes is mediated through enhanced deviant peer associations.

A Dynamic Cascade Model of Development

The aforementioned body of empirical evidence suggests that predictor domains for adolescent violence may intercorrelate, as in Figure 1. Reducing the paths in this figure to the minimum necessary uniquely contributing components is one goal of efficient longitudinal modeling. We propose the dynamic cascade model represented in Figure 2 to describe this developmental process. We recognize that alternate cascade models could hold as well.

This model begins with a child's birth into an adverse social context, characterized by high neighborhood disadvantage and low neighborhood access to social services. It is hypothesized that this context places a child at risk for a variety of problem outcomes throughout life, including chronic antisocial behavior in adolescence. Although a comprehensive model would also include the domain of prenatal experience and temperament, which act as biologically based dispositions to act impulsively and without delay of gratification, the current empirical study began with children at age 5 and could not unambiguously measure this domain. Thus, it is not discussed further.

It is hypothesized that the domain of adverse social context operates on adolescence violence primarily indirectly through impact on life circumstances across childhood, especially early parenting. An adverse social context increases family stress and the rate of maternal depression that makes it more difficult to parent effectively, especially during discipline events. It is hypothesized that harsh and inconsistent discipline practices and lack of parental warmth predict adolescent violence, partially mediate the impact of an adverse social context on later violence outcomes, and increment the prediction of violence outcomes beyond context predictors.

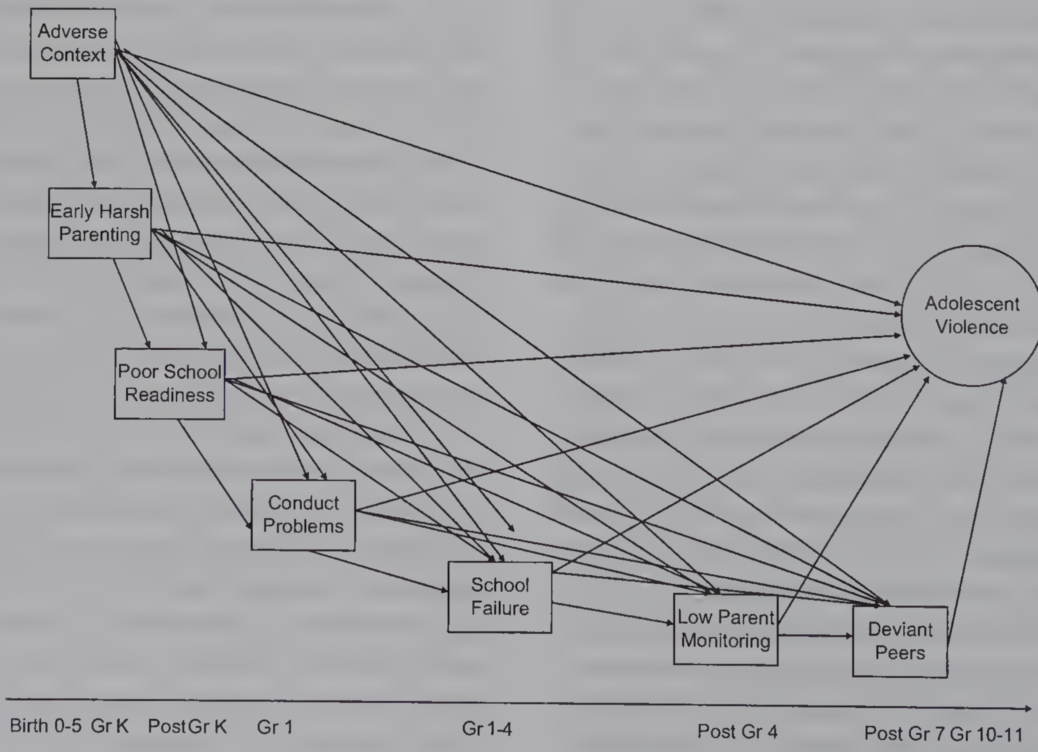


Figure 1. Hypothesized correlations among domains in the development of violence.

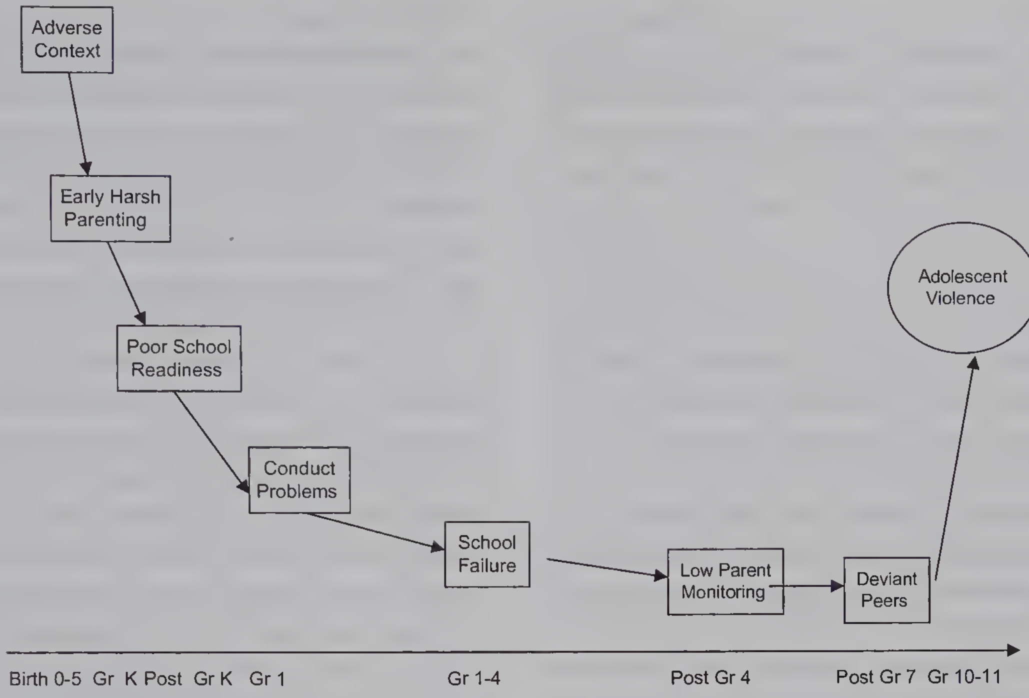


Figure 2. Hypothesized dynamic cascade model of the development of violent behavior.

But how does the impact of early parenting practices operate? These discipline practices are hypothesized to reduce the child's opportunities for learning self-regulatory, social-cognitive, and cognitive skills and attitudes that might be called school readiness skills. Deficits in these skills, in turn, are hypothesized to predict adolescent violence, to mediate the impact of parenting practices on violence outcomes, and to increment the prediction of violence outcomes.

How do school readiness skill deficits, in a context of poverty and harsh and inconsistent parenting, lead to adolescent violence? It is hypothesized that these deficits increase the probability that a child will experience failure in peer relations and academic performance. School social and academic failure, in turn, is hypothesized to predict adolescent violence outcomes, to mediate the impact of school readiness skills deficits, and to increment the prediction of these outcomes beyond the skill deficits themselves.

School failure operates, in turn, partially through its reciprocal impact on the parent–child relationship. School and peer conflict make parenting an even more difficult task that requires intensified monitoring and supervision of a child, enhanced parent–child communication, and enhanced parent involvement at school; however, it is hypothesized that these school problems have a paradoxical effect on lowering parental monitoring, supervision, communication, and involvement, in order to reduce conflict or because the parent has begun to give up on exerting an impact on the child. Low parental monitoring is hypothesized to predict adolescent violence outcomes, to mediate partially the impact of school failure on these outcomes, and to increment the prediction of these outcomes beyond school failure.

Finally, a major product of low parental monitoring and school involvement is the youth's tendency to associate with deviant peers. This pattern is hypothesized to have a proximal impact on potentiating adolescent violent behavior, to mediate partially the impact of low parental monitoring on violent behavior, and to increment the prediction beyond low parental monitoring.

The proposed model is a cascade because the six predictor domains are ordered in temporal sequence with each successive stage posited as a partial product of previous stages but also providing added impact on the process (CPPRG, 1992). It is a dynamic model consistent with that posed by Granic and Patterson (2006) because the steps are not hypothesized to operate only in linear ways at regular time intervals. The model is one in which a high-risk child traverses a deepening stream across development toward a violent outcome, with each stage of development being predicted partially from previous events and providing growing inevitability toward the violent outcome, but also offering a new opportunity to begin a different tributary toward a nonviolent outcome.

Testing the Model

Beyond identifying the major domains of predictors of antisocial behavior and the specific variables within domains, the developmentalist's task is to understand how these processes operate in tandem to produce antisocial outcomes. Rutter and Garmezy

(1983) pioneered the cumulative risk model that posits that prediction is strongest simply by summing the number of risk factors, consistent with the idea that these domains will provide unique increments in predicting youth outcomes. Others have found evidence to support this model (Greenberg, Lengua, Coie, Pinderhughes, & CPPRG, 1999; Sameroff, Bartko, Baldwin, & Baldwin, 1988). However, such a model does not address the intercorrelations among predictors or begin to explain processes through which the risk factors operate. The proposed dynamic cascade model posits that a domain operates on antisocial outcomes by directly influencing the next domain in the hypothesized temporal sequence. This next domain both mediates the impact of the prior domain and catalyzes further antisocial development in an incremental manner.

Although the full dynamic cascade model posits many hypotheses, the following 21 hypotheses were tested in the current study:

> H1a to H1f: Predictor domain x will significantly predict the adolescent violence outcome.
>
> H2a to H2e: Predictor domain x will predict predictor domain $x + 1$.
>
> H3a to H3e: Predictor domain $x + 1$ will mediate (either fully or at least partially) the relation between the previous predictor domain x and the adolescent violent outcome.
>
> H4a to H4e: Predictor domain x will significantly increment the prediction of the adolescent violence outcome beyond the prediction afforded by predictor domain $x - 1$.

Our ability to test hypotheses of the proposed model empirically is limited by the sample being studied, the variables that we measure, and the methods that we apply in analysis. The current study was designed to test key features of this model in a diverse sample of boys and girls from four geographic sites followed from kindergarten through late adolescence. The sample included both the normative and high-risk control samples of the Fast Track Study (CPPRG, 1992), a randomized controlled trial of the efficacy of a preventive intervention designed following the premises of the proposed model. We oversampled, and appropriately weighted, early-starting high-risk children by including the high-risk control children in the randomized trial. No child who was assigned to receive intervention was included in the current study.

The outcome variable was an index of serious violent behaviors in Grades 10 and 11 of high school. This construct is inherently skewed, with only about one fourth of the sample receiving scores greater than 0. Variation above 0 was considered meaningful, however.

The proposed model does not stipulate that variables within a predictor domain represent indicators of a single latent construct defined by that domain. Rather, within-domain variables represent a collection of manifest factors that are conceptually related, measured at roughly the same time point, and possibly (though not necessarily) correlated with each other. They are aggregated in a weighted index, similar to an index of adverse life events or stressors, rather than a latent trait. Thus, we required a data-analytic method that would preserve these characteristics of each predictor domain but yield a single factor for simplicity of hypothesis testing. We chose to compute partial least squares (PLS) variates for each domain and test hypotheses through model contrasts and significance testing by a bootstrap method.

Impact of Gender

Because males are more likely than females to become seriously violent and because of obvious biological and socialization differences across genders, any comprehensive developmental model must contend with the question of whether developmental pathways vary for males and females. Gorman-Smith and Loeber (2005) found similarity with the prior empirical literature on males in the predictors of violence that emerged in their all-female sample. The current sample includes both genders, with sufficient numbers of high-risk females to test major hypotheses in each gender. With the predictor variables studied here, we had no reason to hypothesize gender-specific pathways. Nonetheless, the gender interaction hypothesis was tested by contrasting models that allowed paths to differ (or not) for each gender.

Variation by Level of Initial Risk

Although it is often assumed that predictor variable operates similarly across the range of risk in a population, recent reports of interaction effects between level of genetic or biological risk and environmental risk in predicting antisocial outcomes suggest that the impact of environmental variables might differ for children at high and low early risk. For example, Lynam et al. (2000) found that adverse neighborhood context exerted a stronger effect on antisocial outcomes among impulsive adolescents than nonimpulsive adolescents. Lengua, Wolchik, Sandler, and West (2000) found that maternal inconsistent discipline had a stronger impact on externalizing behavior among impulsive toddlers than among nonimpulsive toddlers. Dodge et al. (2003) found that peer social rejection had a stronger effect on increasing aggressive behavior outcomes among children who had been above the median in aggression initially than children below the median.

Participants in the current study included a community sample of children who represented the normative distribution of risk for antisocial outcomes and a sample of high-risk children selected on the basis of early conduct problems. These samples afforded the opportunity to test whether developmental models varied across risk groups.

Method

Participants

Participants came from the control schools of a longitudinal, multisite investigation of the development and prevention of conduct problems in children, the Fast Track Project (CPPRG, 1992). At each of four sites (Durham, NC; Nashville, TN; Seattle, WA; and central Pennsylvania), high-risk schools were selected and randomly assigned to intervention or control conditions. From among the control schools ($n = 27$), teachers completed ratings of child disruptive behavior to identify a within-site stratified sample of about 10 children within each decile of behavior problems. Across the four sites, 387 children were selected to represent the normative population of these schools. In addition, high-risk children were oversampled to measure finer gradations of high risk. A multistage, multi-informant screening process identified three annual cohorts of kindergartners with highest disruptive behavior scores, yielding an additional n of 367, bringing the total number of participants to 754 (see Lochman & CPPRG, 1995, for further details). Of the normative sample, 35% came from single-parent families, 23% of mothers had not graduated from high school, 50% were male, and 43% were African American. Although the high-risk subsample included higher proportions of each of these characteristics, weighting was used in all analyses to reflect the oversampling of high-risk children.

Procedures

Beginning in kindergarten and lasting through Grade 11, annual measurements were collected from multiple sources that included the teacher, the peer group, administrative school records, the mother, the child, and interviewer ratings. Teacher, peer, and administrative record measures were collected in the spring, and mother, child, and interviewer measures were collected in the following summer.

Measures

The following measures used in the current study are described in detail at www.fasttrackproject.org. All measures were scored so that higher scores are more negative. Prior to administration, interviewers received extensive training, including reading manuals, observing videotapes, participating in guided practice and role plays, observing trainers conduct home visits, completing pilot home visits themselves, and having trainers supervise home visits until the interviewers achieved a minimum of 70% agreement on all ratings. During the spring prior to summer home visits, trainers from each study site attended multiday teaching sessions designed to ensure project-wide standardization. During the summer data collection periods, these trainers participated in regular conference calls to contain site-specific drift (Pinderhughes, Nix, Foster, Jones, & CPPRG, 2001).

Adverse social context. Four variables indexed the child's early adverse social context. First, the interviewer who visited the child's home in the summer after kindergarten completed a four-item assessment of the *neighborhood danger* (*1a*) immediately after the visit. The scale summed responses regarding the safety of the dwelling, the street, and the neighborhood, and the disruptiveness of noise, on 4-point scales ($\alpha = .83$, scale of 4–16). Second, during the initial interview in kindergarten, the mother was asked to respond to each of two items on 10-point scales regarding her satisfaction with neighborhood services for transportation and education. These items were summed to yield a score for *lack of neighborhood services* (*1b*) ($\alpha = .41$, scale of 2–20). Third, during the interview the parent reported the education level and occupation of the child's biological father and biological mother. Following Hollingshead (1975), these four responses were scored and combined to form a single score for *family socioeconomic status* (*1c*), ranging from 8 to 66 ($M = 27.8$, $SD = 13$). Fourth, during the interview after Grade 1, the mother completed the Center for Epidemiological Studies Depression Scale (Radloff, 1977). Each of 20 items was rated on a scale of 0–3, yielding a sum score for *maternal depression* (*1d*) ($\alpha = .89$, scale of 0–60). It was first administered after Grade 1, and it was inferred that it represented a stable construct from early life.

Early harsh and inconsistent parenting. Five scores indexed early harsh parenting. During the first home interview following kindergarten, the interviewer presented the parent with six hypothetical vignettes taken from Dodge et al. (1990), each depicting a parent–child interaction in which the child misbehaves. The parent responded with the behavioral reaction that she would give, which was scored for physical punishment (0 = *not mentioned*, 1 = *mentioned*). The items were averaged to yield a score for *harsh discipline—vignette* (2*a*) (α = .60, scale of 0–1). During the same interview, the parent completed the Conflict Tactics Scale (CTS; Straus, 1979), which asks the parent to report on a 7-point scale the frequency with which she displays various discipline tactic toward the child. A scale of *corporal punishment—CTS* (2*b*) was derived as the mean of three items (spank, spank with object, and hit; α = .79, scale of 0–6). During the same interview, the interviewer engaged in an open-ended interview with the parent about discipline of the child, following the protocol defined by Dodge et al. Following a discussion about how the parent typically disciplined the child, including a discussion of the most severe episodes, the interviewer made a private 5-point scale rating of the harshness of the discipline received by the child, called *harsh discipline—interviewer rating* (2*c*). This measure was adapted from Lansford, Deater-Deckard, Dodge, Bates, and Pettit (2004), who conducted independent coder agreement checks on 56 randomly selected mothers and found that intercoder agreement was acceptable (r = .80, p < .001).

During the same home visit, the parent and child were asked to complete four interaction tasks taken from Eyberg (1974): a child's game, a parent's game, a Lego task, and cleanup. Interviewers were trained in the administration of this task and coding of behaviors by a supervisor with prior experience with this measure or someone trained by that person until they reached acceptable agreement. This well-validated measure has been used by one of the current study authors (Crnic & Greenberg, 1990) in numerous studies. The interviewer observed the interaction and immediately made 5-point ratings, taken from Crnic and Greenberg (1990). Six items were averaged to yield a scale of *lack of parental warmth—observed* (2*d*) (e.g., parent's apparent gratification, sensitivity, and involvement during each segment; α = .90, scale of 1–5). Finally, the parent completed the Parent Questionnaire (Strayhorn & Weidman, 1988), which included seven items asking the parent to report (using a 5-point scale) her use of appropriate and consistent discipline strategies (e.g., "what percent of time do you follow up on your commands?" "How much does your discipline vary with your mood?"). These responses were averaged to yield a score for *inconsistent discipline—parent report* (2*e*) (α = .67, scale of 1–5).

School social and cognitive readiness. Five variables indexed social and cognitive readiness. During the first interview with the child, the child was administered the standardized 57-item Woodcock–Johnson Test, which yielded a score for *Woodcock–Johnson Letter/Word Identification* (3*a*) (α = .88, scale of 0–57). The child also completed the Wechsler Intelligence Scale for Children vocabulary subtest, which yielded a scaled score for *vocabulary skill* (3*b*), which has a nationally normed mean of 10 and an *SD* of 3. The child completed the Social Problem Solving Test, adapted from Dodge et al. (1990), in which the child was presented with each of eight drawings depicting social dilemmas (e.g., "A peer is playing with your bicycle and you want it back; how could you get it back?") and was asked how he or she could solve the dilemma. Responses were recorded verbatim and categorized later as competent or not. The proportion of responses across vignettes categorized as competent constituted a score for *competent social problem solving* (3*c*) (α = .66, scale of 0–1). The child also completed the Home Interview With Child measure (Dodge et al., 1990) which assesses hostile attributional biases. The child responded to each of eight drawings depicting hypothetical provocation vignettes in which the child imagined being provoked by a peer (e.g., "A peer bumps into you.") by stating whether the peer had likely acted with benign or hostile intent. The proportion of times that a child responded with hostile intent was scored as a measure of *hostile attributional bias* (3*d*) (α = .78, scale of 0–1). The child also completed the Emotion Recognition Questionnaire, adapted from Ribordy, Camras, Stefani, and Spaccarelli (1988), which consists of 16 vignettes depicting a child experiencing an emotion (happy, sad, fear, angry). The child was asked to identify the emotion being depicted. The number of correct responses was scored as a measure of *emotion recognition skill* (3*e*) (α = .65, scale of 0–16).

Child externalizing problems. The child's externalizing behavior problems were assessed using the Child Behavior Checklist (Achenbach, 1991) with parent and teacher reports collected at the end of kindergarten. The highly reliable *T* scores for the 50-item externalizing problems scale were log transformed to produce scores for *child externalizing problems—teacher report* (4*a*) and *child externalizing problems—parent report* (4*b*).

School social and academic failure. Five scores indexed elementary school social and academic failure.

During the spring of Grade 2, all consenting peers (at least 75% of parents in each classroom for which scores are tabulated provided written informed consent) in a child's classroom were asked to complete a sociometric interview in which they were asked to nominate children whom they like and whom they dislike. The numbers of nominations received in each category were summed, standardized within classroom, and differenced. A score for *peer social rejection* (*5a*) (0 = *not rejected*, 1 = *rejected*) was computed following the formula by Coie, Dodge, and Coppotelli (1982). *Grade retention* (*5b*) was scored from the child's official school records in Grades 2, 3, and 4 as 0 (*never retained*) or 1 (*ever retained in grade*). Placement into *special education* (*5c*) was scored from the same records as 0 (*never*) or 1 (*ever*). Core subject *failing grades* (*5d*) was scored from the same records as whether the child had ever received a failing grade in a core subject, scored as 0 (*never failed*) or 1 (*ever failed*). Finally, *academic competency problems* (*5e*) was scored based on the child's performance on the three scales of the 158-item Woodcock–Johnson Test in Grade 4 (α = .87). A child was given a score of 0 if he or she received a score lower than 1.5 *SD*s below the sample mean on any of three subtests (passage comprehension, letter/word identification, and calculation), and 1 otherwise.

Adolescent parental monitoring. Six scales were collected from parents and children during or after Grade 4 to assess parental monitoring and supervision of the child. The first scale, called *parental supervision—child report* (*6a*) was derived as the mean of five items of the Supervision/Involvement Scale of the Pittsburgh Youth Study (Loeber, Farrington, Stouthamer-Loeber, & von Kammen, 1998). For each item, the youth responded no or yes (e.g., "Does your parent know who you are with when you are away from home?"; α = .62, scale of 1–5). The second scale, called *parental supervision—parent report* (*6b*), was derived from the same source and asked parents the same items from their perspective (α = .55, scale of 1–5). The third scale, adapted from the Revised Parent–Adolescent Communication Form of the Pittsburgh Youth Study (see Loeber et al., 1998), includes five items (coded on a 5-point scale from *almost never* to *almost always*) that assess a child's perceptions of his or her *parent openness to communication* (*6c*) (e.g., "My parent is a good listener."; α = .56, scale of 1–5). The fourth scale was adapted from the same source as the third scale and includes five items that assess the parent's perceptions of his or her *parental communication with child* (*6d*) (e.g., "I discuss my child's problems with him or her.") (α = .73, scale of 1–5). The fifth scale was adapted from the same source as the third scale and includes five items that assess the child's perceptions of his or her *communication with parents* (*6e*) (e.g., "I discuss my problems with my parents"; α = .75, scale of 1–5). The sixth scale, developed for this study, is the *parent involvement in school* (*6f*), for which the teacher answered twenty-one 5-point-scale items about the parent's contact, involvement, and comfort with the teacher (α = .91, scale of 0–4).

Adolescent deviant peer associations. Two scales indexed this domain after Grade 7. For the first scale, the parent completed the Parent Report on Child's Close Friends, consisting of 10 questions developed for the current study that assessed *deviant peer associations—parent report* (*7a*) (α = .87, scale of 4–40), centered around three main topics: the child's best friend, the child's second best friend, and the child's close friends. For each topic, the parent was asked to respond on a 4-point scale to these questions (a) Does this friend get into trouble with teachers, the police, or other adults? (b) Does this friend do things you disapprove of? (c) Are you concerned about the negative influence this friend has on your child? The tenth 5-point scale question asked parents, "How much does your son/daughter hang around with kids who get into trouble?" For the second scale, the youth completed the Self-Report of Close Friends, which included five items developed for the Seattle Social Development Project (O'Donnell, Hawkins, & Abbott, 1995) that measured *deviant peer associations—youth report* (*7b*) (α = .82, scale of 1–4). The youth was asked to identify two best friends and answer questions about behaviors by the friend including delinquency, substance use, school motivation, and parental attitudes toward the friend.

Adolescent violent behavior. In the summers after Grades 10 and 11, the youth and parent reported the youth's violent behavior in the past 12 months. Six youth-report items each year were taken from Huizinga and Elliot's (1987) Self-Report of Delinquency, and eight items each year were taken from the Guns and Gangs instrument developed for this study (CPPRG, 1999). Two parent-report items each year were taken from the Parent Report of Delinquency (CPPRG, 1999). The 32 items (listed in Table 1) were each scored as 0 if *no* and 1 if *yes*, with the sum indexing *serious interpersonal violence* (*8*) (α = .49, scale of 0–16).

Modeling Strategy

We developed a two-stage modeling process to accommodate the distinctive characteristics of our measurement model and to evaluate the structural model properly. First, we applied PLS (Chin, 1995) modeling to develop the measurement model for

Table 1
Variables Used in Index of Adolescent Violent Behavior

No.	Variable	Source	Wording
	Self-report of delinquency		
1	C11AD26a	Self	In the past year, have you attacked anyone with the intent to hurt or kill?
2	C11AD27a	Self	In the past year, have you hit someone with the idea of hurting them?
3	C11AD28a	Self	In the past year, have you used a weapon to get money from people?
4	C11AD29a	Self	In the past year, have you thrown objects at people?
5	C11AD30a	Self	In the past year, have you been involved in any gang fights?
6	C11AD32a	Self	In the past year, have you had sex with someone against their will?
	Guns and gangs		
7	C11AW2d	Self	In the past 12 months, have you ever pulled a small pocket knife on another person?
8	C11AW2e	Self	In the past 12 months, have you ever hurt someone with a small pocket knife?
9	C11AW3d	Self	In the past 12 months, have you ever pulled a switchblade, razor, or other big knife on another person?
10	C11AW3e	Self	In the past 12 months, have you ever hurt someone with a switchblade, razor, or other big knife?
11	C11AW4e	Self	Guns in your home that your parents own for their protection or for hunting—In the past 12 months, have you ever pulled it on another person?
12	C11AW4f	Self	Guns in your home that your parents own for their protection or for hunting—In the past 12 months, have you ever hurt some one with it?
13	C11AW5d	Self	A gun that is not for hunting—In the past 12 months, have you ever pulled one on another person?
14	C11AW5e	Self	A gun that is not for hunting—In the past 12 months, have you ever hurt someone with it?
	Parents' report of delinquency		
15	P11AG11	Parent	How often has your child attacked someone with a weapon with the idea of seriously hurting them?
16	P11AG16	Parent	How often has your child carried a weapon?

predictor constructs. PLS modeling as applied here is superficially similar to confirmatory factor analysis (CFA) but uses an alternative estimation technique (vs. maximum likelihood) that has its own advantages and disadvantages. In PLS, summary variables are modeled with formative indicators, versus reflective indicators usual for latent variables in CFA. It is also possible to model variables with formative indicators in structural equation modeling (SEM), but the requirements for model identification are much more stringent. Reflective approaches are appropriate when component variables are presumed to be indicators of a single latent construct (such as a trait), whereas formative approaches make no such assumption (such as indicators of a total stress index or life events index). The formative indicator approach involves each summary variable being modeled as a linear composite of its components, as in a principal components analysis, rather than being a source of common variance among a factor's indicators. The key difference between a PLS analysis and a principal components analysis is that in the latter, the composites are constructed to maximize the variance of the indicators accounted for by the component, whereas in PLS, the indicators for each composite are weighted to maximize the strength of the relations among the PLS variates (similar to a canonical correlation), resulting in the best possible prediction of the downstream variate. Rather than assume equal weights for all indicators of a scale, the PLS algorithm allows each indicator to vary in how much it contributes to the composite score. Thus, indicators with weaker relations to indicators and the construct are given lower weightings. In this sense, PLS is preferable to techniques such as regression that assume error-free measurement (Lohmoller, 1989; Wold, 1982).

Once the PLS weights for the predictors were estimated, we tested our hypotheses about the relations among these blocks of variables in an SEM framework, using standardized scores for the variates to ease interpretation of results. Given our intent in predicting the interpersonal violence outcome, we estimated seven separate PLS models. In each model, one time block of predictors (social context, early parenting, etc.) formed the PLS variate, weighted to maximize the prediction of the violence score. An alternative strategy would have been to estimate weights for all seven variates in a single PLS model; however, the PLS algorithm would have then also worked to maximize prediction among the seven predictor variates, at the cost of maximal prediction of interpersonal violence. We then retained the individual scores. Weights for the scores are shown in Table 2.

Results

Descriptive Statistics and Treatment of Missing Data

Descriptive statistics (means, standard deviations, missing data rates, and correlations with the adolescent violence score) for the manifest variables used in the analysis, incorporating the sample weights reflecting the oversampling of higher risk students, are shown in Table 2. We conducted our PLS modeling in SAS Version 9.1.3 (SAS Institute, 2004), which includes the option to accommodate missing data by estimating

Table 2

Means, Standard Deviations, Percent Missing, Partial Least Squares (PLS) Variate Weights, and Correlations With Violence Outcome

Variable	*M*	*SD*	% missing	PLS	*r* w/viol
Early adverse social context					
1a. Neighborhood danger (age 0–5)	8.47	2.37	4.1	.080	.08
1b. Lack of neighborhood services (age 0–5)	14.92	4.35	0.4	.144	.25*
1c. Family socioeconomic status (age 0–5) (R)	37.80	54.60	0.0	−.077	−.14*
1d. Maternal depression (age 6)	11.97	9.39	0.0	.128	.18*
Early harsh and inconsistent parenting					
2a. Harsh discipline—vignette (ages 1–5)	0.19	0.21	0.2	.042	.03
2b. Corporal punishment—CTS (ages 1–5)	1.30	1.30	0.0	.074	.16*
2c. Harsh discipline—interviewer rating (ages 1–5)	2.20	0.80	0.0	.061	.05
2d. Lack of parental warmth—observed (ages 1–5)	3.86	0.74	2.2	.020	.12*
2e. Inconsistent discipline—parent report (ages 1–5)	2.82	0.55	1.3	.076	.15*
School social and cognitive readiness					
3a. Woodcock–Johnson Letter/Word Identification (Kindergarten) (R)	13.83	4.68	0.0	.065	.11*
3b. Vocabulary skill (Kindergarten) (R)	9.41	3.55	0.0	.028	.00
3c. Competent social problem solving (Kindergarten) (R)	0.72	0.16	0.6	.011	.08
3d. Hostile attributional bias (Kindergarten)	0.59	0.30	0.0	.089	.22*
3e. Emotion recognition skill (Kindergarten) (R)	11.48	2.74	0.9	.078	.10*
Early child externalizing behaviors					
4a Teacher report—TRF (log) (Grade x)	4.18	0.44	66.5[a]	.047	.07
4b. Parent report—CBCL (log) (Grade x)	3.94	0.82	6.9	.015	.25*
School social and academic failure					
5a. Peer social rejection (Grades 2–4)	0.15	0.89	22.7	.118	.21*
5b. Grade retention (Grades 1–4)	0.11	1.29	0.0	.034	.00
5c. Special education (Grades 1–4)	0.24	1.77	0.0	.072	.06
5d. Failing grades (Grades 1–4)	0.06	0.98	0.0	.031	.00
5e. Academic competency problems (Grades 1–4)	0.23	1.75	0.0	.044	.03
Adolescent parental monitoring					
6a. Parental supervision—youth report (after Grade 4) (R)	4.22	1.10	10.4	.073	.09
6b. Parental supervision—parent report (after Grade 4) (R)	4.79	1.15	12.3	.047	.06
6c. Parent openness to communication (after Grade 4) (R)	1.84	0.76	14.5	.074	.11
6d. Parental communication with child (after Grade 4) (R)	3.88	0.65	9.3	.062	.15*
6e. Communication with parents (after Grade 4) (R)	3.72	1.00	11.9	.021	.03
6f. Parent involvement in school (after Grade 4) (R)	1.42	0.58	11.9	.059	−.04
Adolescent deviant peer association					
7a. Deviant peer association—parent report (after Grade 7)	15.23	5.62	18.1	.143	.27*
7b. Deviant peer association—youth report (after Grade 7)	1.34	0.60	21.4	.114	.14*
Index of adolescent violence (Grades 10–11)	0.54	6.21	19.2		

Note. n ranges from 280 to 462. w/viol = with violence; CTS = Conflict Tactics Scale; TRF = Teacher Rating Form; CBCL = Child Behavior Checklist.

[a] Not administered to normative sample.

*$p < .05$.

values for the scores through the EM algorithm. One important component of this modeling is estimation of the standard error in the context of missing data estimation, which we completed in Mplus Version 4.1 (Muthén & Muthén, 2006), using its facility for the bootstrap, and accommodating the sample weights. Standard errors for this and all further analysis were based on simple bootstrap of 500 samples. The combination of the EM algorithm and bootstrapped standard errors adjusts for missing data under the assumption of missing at random (MAR). The MAR assumption is that all relations between the probability of a response being missing and the true value of that response are accounted for by other observed variables in the model. The MAR assumption is not logically testable; however, analyzing data that do not meet this assumption by MAR methods has been shown to be preferable to ad hoc approaches to missing data such as listwise deletion (Graham, Hofer, & MacKinnon, 1996). We estimated the correlations among the PLS variates (and the interpersonal violence score) in Mplus v.4.1 (Muthén & Muthén, 2006). Correlations appear in Table 3.

We then tested the cascade nature of the relations among PLS variates. It was hypothesized that each variate would cascade to predict the next variate immediately downstream. We tested each variate (except the first social context) as a mediator of the relation between the immediately prior variate and the ultimate outcome, interpersonal violence. Mediation was tested by the comparison of the product of the two relevant coefficients (prior to central, central to posterior) to its bootstrapped standard error, using significance thresholds determined and recommended by MacKinnon, Lockwood, Hoffman, West, and Sheets (2002, 2006). The coefficient is unbiased in testing the statistical significance of at least partial mediation. Whether mediation is full or partial is tested by the coefficient for the predictor variate in predicting the outcome net of the effect of the mediator. All relevant statistics appear in Table 4.

Results of Hypothesis Tests

Adverse social context. Hypothesis H1a was supported: Adverse social context was a significant predictor of adolescent violence, $r = .30$. Hypothesis H2a was supported: Adverse social context was

Table 3

Correlations Among Partial Least Squares Variates and Violence

	Adverse context	Early harsh parenting	Social/cognitive deficits	Externalizing behavior	Social/academic failure	Low monitoring	Deviant peer
Early harsh parenting	.42*						
Social and cognitive deficits	.34*	.33*					
Externalizing behavior	.36*	.36*	.12*				
School social and academic failure	.22*	.28*	.30*	.18*			
Low parent monitoring	.28*	.35*	.26*	.26*	.23*		
Deviant peer association	.31*	.22*	.20*	.25*	.15*	.20*	
Violence	.30*	.20*	.20*	.25*	.18*	.16*	.26*

Note. $N = 463$.
*$p < .05$.

Table 4

Tests of Direct and Indirect Effects in Predicting Violence Outcome

Source	Adverse context	Early parenting	Social/cognitive readiness	Externalizing behavior	School failure	Parental monitoring
Mediator	Early parenting	Social/cognitive readiness	Externalizing behavior	School failure	Parental monitoring	Deviant peer association
H1: Source → violence	.301 (.077)*	.182 (.044)*	.215 (.044)*	.277 (.073)*	.163 (.041)*	.152 (.044)*
H2: Source → mediator (α)	.384 (.040)*	.382 (.049)*	.122 (.046)*	.147 (.034)*	.234 (.046)*	.198 (.050)*
H4: Mediator → violence (β)	.099 (.044)*	.147 (.037)*	.213 (.058)*	.158 (.048)*	.139 (.050)*	.272 (.087)*
H3: Indirect effect ($\alpha\beta$)	.014 (.017)*	.056 (.015)*	.026 (.013)*	.023 (.009)*	.033 (.013)*	.054 (.023)*

Note. "Source" indicates the predictor variable. Figure is coefficient. Standard error is in parentheses. $N = 463$.
*$p < .05$.

a significant predictor of harsh and inconsistent early parenting, $r = .42$.

Harsh and inconsistent early parenting. Hypothesis H1b was supported: Harsh and inconsistent early parenting was a significant predictor of adolescent violence, $r = .20$. Hypothesis H2b was supported: Harsh early parenting was a significant predictor of lack of school readiness, $r = .33$. Hypothesis H3a was supported: Harsh and inconsistent early parenting partially mediated the prediction of adolescent violence from adverse social context, *Est./SE* = 2.21. Hypothesis H4a was supported: Harsh and inconsistent early parenting significantly incrementally predicted adolescent violence, controlling for adverse social context, $b = .099$.

Lack of school social and cognitive readiness. Hypothesis H1c was supported: Lack of school readiness was a significant predictor of adolescent violence, $r = .20$. Hypothesis H2c was supported: Lack of school readiness was a significant predictor of early child externalizing problems, $r = .12$. Hypothesis H3b was supported: Lack of school readiness partially mediated the prediction of adolescent violence from harsh early parenting, *Est./SE* = 3.75. Hypothesis H4b was supported: Lack of school readiness significantly incrementally predicted adolescent violence, controlling for harsh and inconsistent early parenting, $b = .147$.

Early child externalizing problems. Hypothesis H1d was supported: Early externalizing problems were a significant predictor of adolescent violence, $r = .25$. Hypothesis H2d was supported: Early externalizing problems were a significant predictor of school failure, $r = .18$. Hypothesis H3c was supported: Early externalizing problems partially mediated the prediction of adolescent violence from lack of school readiness, *Est./SE* = 2.01. Hypothesis H4c was supported: Early externalizing problems significantly incrementally predicted adolescent violence, controlling for lack of school social and cognitive readiness, $b = .213$.

School social and academic failure. Hypothesis H1e was supported: School failure was a significant predictor of adolescent violence, $r = .18$. Hypothesis H2e was supported: School failure was a significant predictor of lack of parental monitoring, $r = .22$. Hypothesis H3d was supported: School failure partially mediated the prediction of adolescent violence from early externalizing problems, *Est./SE* = 2.61. Hypothesis H4d was supported: School failure significantly incrementally predicted adolescent violence, controlling for childhood externalizing behavior problems, $b = .158$.

Lack of parental monitoring. Hypothesis H1f was supported: Lack of parental monitoring was a significant predictor of adolescent violence, $r = .16$. Hypothesis H2f was supported: Lack of parental monitoring was a significant predictor of deviant peer involvement, $r = .20$. Hypothesis H3e was supported: Lack of parental monitoring partially mediated the prediction of adolescent violence from school failure, *Est./SE* = 2.53. Hypothesis H4e was supported: Lack of parental monitoring significantly incrementally predicted adolescent violence, controlling for school failure, $b = .139$.

Deviant peer associations. Hypothesis H1g was supported: Deviant peer involvement was a significant predictor of adolescent violence, $r = .26$. Hypothesis H3f was supported: Deviant peer involvement partially mediated the prediction of adolescent violence from lack of parental monitoring, *Est./SE* = 2.30. Hypothesis H4f was supported: Deviant peer involvement significantly incrementally predicted adolescent violence, controlling for lack of parental monitoring, $b = .272$.

Tests of Full Model

The next stage of analysis was to construct a full path model showing the prediction of each construct from all prior constructs. In this model, estimated using Mplus Version 4.2 (Muthén & Muthén, 2006), each variate was incorporated predicting each subsequent variate and interpersonal violence. Parameter estimates and bootstrapped standard errors are shown in Table 5. This overall model was saturated and thus not testable. Tests of paths, however, indicated that each step was significantly predicted from the previous steps taken together, and the best prediction of violence came from multiple steps.

Gender Differences in Means and Developmental Paths

Simple *t* tests indicated that males received significantly higher raw adolescent violence mean scores than females ($p < .001$). Males and females did not differ in the adversity of their home context, in early parenting, or in early externalizing problems scores (only because *T* scores that standardize within gender were used), but males received significantly higher scores than females for lack of school social and cognitive readiness ($p < .02$), school social and academic failure ($p < .001$), lack of parental monitoring and communication ($p < .02$), and deviant peer associations ($p < .001$).

To test the hypothesis that developmental paths would differ across gender groups, the full model was contrasted with a model in which path coefficients were allowed to vary across male and female groups. Loosening the constraint that the paths are identical did

Table 5

Prediction Between Partial Least Squares Variates and Violence in Full Model

	Step 1: Adverse context	Step 2: Early harsh parenting	Step 3: Social/cognitive readiness	Step 4: Early externalizing problems	Step 5: School social/academic failure	Step 6: Adolescent parental monitoring	Step 7: Deviant peer association
Step 2:Eearly harsh parenting	.384 (.040)*						
Step 3: Social/cognitive readiness	.262 (.048)*	.264 (.049)*					
Step 4:Eearly externalizing problems	.294 (.051)*	.331 (.055)*	−.062 (.046)				
Step 5: School social/academic failure	.042 (.045)	.159 (.056)*	.191 (.044)*	.065 (.037)			
Step 6: Adolescent parental monitoring	.079 (.047)	.216 (.053)*	.108 (.041)*	.106 (.035)*	.093 (.048)		
Step 7: Deviant peer association	.187 (.048)*	.032 (.060)	.068 (.038)*	.117 (.043)*	.029 (.048)	.071 (.050)	
Violence	.174 (.066)*	.000 (.066)	.071 (.031)*	.116 (.048)*	.083 (.047)	.017 (.057)	.171 (.075)*

Note. $N = 463$. Figure is path coefficient. Standard error is in parentheses.
$^*p < .05$.

significantly increase model fit, $\chi^2(28, n = 462) = 55.73$, $p = .001$, indicating that the strength of developmental paths differed for males and females. In follow-up analyses, we probed this interaction in each of seven models allowing the paths to differ at each step in the cascade. A significant interaction with gender was found for one of the seven models, predicting school failure, $\chi^2(4, N = 462) = 18.48, p = .001$.

This model predicting school failure freed across gender groups fit the data well, $\chi^2(24, N = 462) = 37.25$, $p = .041$, comparative fit index [CFI] = .976, Tucker-Lewis index [TLI] = .943, estimated root mean square error of approximation [RMSEA] = .049 (90% confidence interval = .010 – .078). For girls, school failure was significantly predicted by social context, $b = .164$, $SE = .067$, $Est./SE = 2.46$, $p < .05$, and by early parenting problems, $b = .196$, $SE = .082$, $Est./SE = 2.39$, $p < .05$, but not by lack of school readiness, $Est./SE = 1.24$, or by early externalizing problems, $Est./SE = -0.84$. The opposite held true for boys, with significant prediction from early externalizing problems, $b = .189$, $SE = .046$, $Est./SE = 4.15$, $p < .001$, and lack of school readiness, $b = .308$, $SE = .054$, $Est./SE = 5.73$, $p < .001$, but not by social context, $Est./SE = -0.91$, or early parenting problems, $Est./SE = 1.22$.

Tests of Interaction With Level of Initial Risk

The full model was contrasted with a model in which path coefficients were allowed to vary between the high-risk and normative groups. Loosening the constraint that the paths are identical significantly improved model fit, $\chi^2(28, n = 462) = 112.86, p < .001$, indicating that the strength of developmental paths differed for high-risk and normative groups. Of the seven domains predicting violence outcomes, one (deviant peer associations) yielded a significantly stronger path in the normative than in the high-risk group ($p < .05$).

Among the other 15 path coefficients tested and depicted in Table 5, 3 yielded a significant interaction involving level of initial risk. All 3 involved early harsh and inconsistent parenting. The relations between early harsh and inconsistent parenting and school social and cognitive readiness, parental monitoring, and deviant peer associations were stronger in the normative group than in the high-risk group.

Discussion

This study reports four major empirical findings that, together, support a dynamic cascade model of how serious violence develops across the life span from early childhood through adolescence. The major contribution is an understanding of how predictor domains operate in concert to predict violent outcomes. Findings indicate that (a) violent outcomes are multiply predicted from factors in diverse domains spanning early development, (b) these domains are intercorrelated and each domain predicts the next domain in development across childhood, (c) each succeeding domain partially mediates the effect of the immediately precedent domain, and (d) although domains intercorrelate and mediate each other, domains play unique and incremental roles in predicting later violent behavior.

Toward a Dynamic Cascade Model of the Development of Violent Behavior

Each of the seven predictor domains (early adverse context, harsh and inconsistent discipline, school social and cognitive readiness, early externalizing behavior problems, school social and cognitive failure, parental monitoring and communication, and deviant peer associations) was significantly related to adolescent violence. Consistent with the literature reviewed earlier, these findings also extend the literature to the relatively severe outcome of violent behavior and generalize these patterns across both male and female groups. The field's confidence in these domains as risk indicators is enhanced by these findings.

The current study provides the most detailed description to date of how these seven temporally distinct predictor domains are intercorrelated. All temporally adjacent domains were significantly related to each other, suggesting that they might provide partially redundant information in predicting violence outcomes. Indeed, each succeeding domain partially mediated the impact of the prior domain on violence outcomes. The redundancy is only partial, however. Each succeeding domain not only mediated the preceding domain, it also provided a significant increment beyond the previous domain in enhancing the prediction of violence outcomes.

This patterned sequencing is consistent with dynamic systems models (e.g., Granic & Patterson, 2006), and it provides a coherent developmental story of how violent behavior grows across childhood and adolescence in a dynamic cascade. The story involves reciprocal impact of parenting on peer relations and peer relations on subsequent parenting. It also involves reciprocal influences between the self and the environment. It explains how early risk is realized across time but also how each new developmental era affords new risk.

The theoretical model that has been supported here starts with children who are born into an adverse neighborhood or disadvantaged family context. This context places children at risk for adolescent violence partly because features of the context (perhaps stress, lack of resources, lack of support; McLoyd, 1990) make it difficult for parents to avoid harsh and inconsistent discipline with their young child. These harsh discipline styles not only account for part of the effect of adverse context on outcomes but they also enhance risk for violence outcomes.

In turn, a partial mechanism through which harsh and inconsistent discipline operates on later violence outcomes is by preventing the child from acquiring social and cognitive skills that are necessary for school social and academic success. These skill deficits include vocabulary deficits, poor social problem solving, hostile attributional biases, and emotion recognition deficits. They partly account for how early dysfunctional parenting leads to adolescent violence. These deficits are more proximal to the phenomenon of adolescent violence (both temporally and theoretically), and they enhance children's risk beyond that afforded by adverse contexts and early parenting. They also predict the child's display of externalizing conduct problems at the time of school entry, which becomes the path leading toward later adverse outcomes.

Externalizing conduct problems at school entry signals the "early starter" in Moffitt's (1993) model of life-course-persistent antisocial behavior and is an important part of the proposed cascade. Methodologically, it is important that this domain is included in the model because a primary alternative hypothesis to the proposed cascade is simply within-child consistency of behavior (from early conduct problems to later violence). Inclusion of this domain enabled testing of this alternative hypothesis, which was empirically rejected by tests that indicated that other domains provide significant increments in the prediction of adolescent violence beyond the prediction afforded by early conduct problems.

Next in the cascade is the role of school social and academic failure, which follows from early externalizing conduct problems. Not surprisingly, conduct problems lead to peer social rejection and academic failure, which account for some of the effect of conduct problems on long-term outcomes. Failure experiences with peers and academic subjects enhance risk for violence in adolescence and increase the likelihood of problem outcomes.

According to the theoretical model, one mechanism through which childhood school failure leads to adolescent violence is through its impact on the parent–child relationship and interaction during early adolescence. Just at a time when youths need monitoring, supervision, and communication with parents, the youths who are at greatest risk suffer from the least amount and poorest quality parenting in these tasks. Parenting during early adolescence not only mediates the impact of the child's previous school failure, it enhances the prediction of adolescent violence. Finally, one way that parental lack of monitoring and communication with a youth leads that youth toward violence is by allowing the youth to interact with deviant peers. Deviant peer associations are the most proximal factor in adolescent violence. They partially mediate the impact of poor parental monitoring, and they enhance this risk as well.

The reciprocal influences between the child and parent over time are especially noteworthy. The parent's early discipline strategies have an impact on the child's social and cognitive readiness for school, which affect conduct problem behavior. In turn, the child's experiences at school cycle back to exert an impact on parenting behavior in early adolescence, which then affects the youth's interactions with peers during adolescence. Bidirectional relations between domains of parenting and child–peer functioning across development have been hypothesized by transactional theorists such as Sameroff and Mackenzie (2003) but rarely have they been empirically demonstrated in longitudinal inquiry.

An important implication of the current findings is that it is premature to conclude that an early-starting antisocial 5-year-old is unequivocally destined for a life-persistent path toward violent outcomes. Although the risk is substantial, it is by no means certain. The findings reported here indicate that trajectories can be deflected at each subsequent era in development, through interactions with peers, school, and parents along the way. The theory of the early-starting antisocial child must account for the deflections in trajectories (e.g., positive outcomes in the face of early risk factors) that have been observed here.

Opportunities for Preventive Intervention

The reciprocal side of the finding that each new developmental era provides enhanced risk for violent outcomes is that each era also provides an opportunity for a change in the youth's developmental trajectory. Even though a child may be at risk due to being born into an adverse neighborhood context, that risk can be lessened by subsequent life events, most proximally the quality of parenting that the child receives during early childhood, especially during discipline encounters. In turn, even though the experience of harsh and inconsistent parenting places a child at risk for violent outcomes, that risk can be lessened by success at school in cognitive and social realms. Likewise, a difficult early school experience that places a child at risk for adolescent violence can be partially offset by parents who closely monitor and supervise their child and communicate with him or her. Finally, risk that is raised by parents who fail to monitor can be lessened if the youth's exposure to deviant peers is limited.

The implications for prevention are many. First, each era affords a new opportunity for targeted intervention in a specific domain that had not been encountered previously. The current findings provide specific targets for prevention at specific eras in development (e.g., during preschool, parenting interventions focused on discipline strategies should be targeted; during the early school years, child-focused intervention directed at enhancing cognitive and social skills and school success should be emphasized) . Second, because new risks arise with each new developmental era, preventive intervention cannot be deemed "over" until the child has passed through adolescence. Third, it appears that the most effective preventive interventions will address multiple aspects of a child's family and school life across multiple eras of development. It is not surprising that, indeed, the most effective preventive intervention programs comprehensively address multiple domains (e.g., CPPRG, 2004; Dishion & Kavanagh, 2000; Henggeler, Schoenwald, Bourduin, Rowland, & Cunningham, 1998; Tremblay, Pagani-Kurtz, Masse, Vitaro, & Pihl, 1995).

Contribution to Knowledge of Antisocial Development in Girls

Very few studies have included sufficient numbers of high-risk girls to make conclusions about developmental pathways toward violence in girls. The findings support the hypothesis that violent outcomes in girls follow largely similar developmental pathways as those for boys. Only one significant difference in pathways was detected, indicating that school failure was more a function of adverse context and parenting for girls than boys, whereas school failure was more a function of externalizing problems for boys than girls. Context, parenting, skills, and peer relations all operate to predict antisocial outcomes in girls in the same way that they do in boys.

Contribution to Methods in Longitudinal Inquiry

Sameroff and Mackenzie (2003) lamented the dearth of longitudinal studies and sophisticated methods that are available to test complicated hypotheses of transactional effects and reciprocal influences across time. Dodge, Coie, et al. (2006) cautioned that the diverse array of variables that have been correlated with antisocial outcomes beg for an integrated theory. The current study offers a novel approach to combining variables in a coherent manner that goes well beyond the simple (though powerful) approach of cumulating risk factors in a summative fashion. The approach requires clearly articulated hypotheses, data collected across multiple nonoverlapping time points, and a sufficient sample size. Variables have been combined into "domains" that represent coherent sets, but the method of combining variables differs critically from the traditional approach of latent variables analyses and SEM, which assume that variables are indicators that are "caused" by a latent construct. Instead, the current approach assumes that manifest variables can be aggregated into an "index" without assumptions of a single latent construct. The data-analytic methods are not new but are newly applied to social development. Dodge, Malone, et al. (2006) have applied similar methods with different domains of variables to understand how children grow into users of illicit substances.

Findings That Provide a Basis for Future Inquiry

Two findings provide the basis for further inquiry. First, findings of partial, rather than full, mediation indicate that precedent stages continue to predict later violence. It might be that risk induced by a prior stage imparts risk that endures in spite of subsequent events, or it might be that the current study simply failed to identify the processes that fully mediate later violence outcomes. Future studies should test hypotheses about other developmental pathways at each stage that might mediate violence outcomes. These other pathways might involve new variables within the same domains that were tested here (e.g., executive cognitive functions such as inhibitory control might be added to the cognitive readiness domain), or they could assess new domains altogether (e.g., quality of preschool experience, afterschool contexts). Furthermore, the categorization of mediation decisions into "full" or "partial" or "no" groups ignores quantitative differences in the magnitude of mediation. Future studies might employ alternate methods of quantifying mediation, using proportion of variance measures or new methods.

Second, it was found that the strength of factors in developmental pathways differed according to level of initial risk. Although not numerous or large, these findings suggest that future inquiry might focus on models of development that apply specifically to high-risk youth.

Finally, a third focus of future inquiry might be to address possible interaction effects among risk domains. That is, it is plausible, even likely, that the impact of one domain is altered (e.g., exacerbated, magnified, mitigated, muted) by variables in another domain. These hypotheses would be tested by interaction effects analyses that were not conducted here.

Limits of the Study

The limits of this study are substantial. They begin with the inescapable fact that these data, although longitudinal, are correlational and are subject to alternate interpretations of third-variable causation, selection factors, and spuriousness of relations. It is still plausible that child factors lead to each of the environmental factors measured here, although the unique increments afforded by different environmental variables suggests that multiple child factors would likely be involved. It is also plausible that each environmental factor is merely a proxy for an unmeasured other environmental factor that causes both the measured variable and its downstream correlates. Further, it may be the some effects that are ascribed to parenting or other environmental factors may be partially the result of genetic effects or gene X environment interaction. No data-analytic method can escape this possibility, but only randomized intervention experiments come closer to parsing cause from spuriousness.

A second limit is that these analyses do not rule out other plausible cascade models that include additional domains and variables. The findings reported here support a cascade model over a simple continuity-of-behavior model, but future studies will need to test competing cascade models against each other. A particular limit is that measurement began when children were 5 years old; thus, reciprocal relations among adverse context, parent characteristics, and very early behavior problems have not been tested. The current model is *a* cascade model, not necessarily *the only* cascade model.

The PLS modeling applied here has been subject to criticism that it makes assumptions about the aggregation of variables into index scores that operate as a unitary whole (Chin, 1995). This aggregation method departs from traditional latent variables approaches and is theoretically defensible in cases when the constituent variables clearly do not represent a unidimensional construct but have heuristic (and temporal) coherence nonetheless. A second concern about the PLS approach is that it may capitalize on chance through its optimization of weighting variables in predicting the downstream outcome. This criticism is the same at that cast toward canonical correlation. The PLS approach is indeed subject to that criticism in the tests of prediction of violent outcomes; however, the abundance of significant bivariate correlations between predictors and the violence outcome in the current study suggest that the conclusions made are not biased by the PLS approach. Furthermore, the tests of relations across PLS variates and the tests of mediation are not at all biased in this manner. They are as rigorous as tests in a structural equation model framework.

The complexity of the dynamic cascade theory suggests methods of data analysis that have not been followed in the current study, such as analyses of accelerating change and nonlinear relations among variables and person-centered approaches to testing hypotheses. The theory also suggests that numerous other variables and other points in time might have been selected for analysis instead of the ones that were chosen. The sequencing of domains for analysis required that any variable be located within a single time point, but realities of social development indicate that these same variables could be located at multiple time points. Furthermore, the arbitrariness of the unit of time as 1 year is obviously flawed. Dynamic development, transactional exchanges, and reciprocal influences undoubtedly occur within shorter time intervals than those measured here. Although the findings support the transactional nature of the relationship between child and parent, undoubtedly these transactions are far more frequent and powerful than could be captured by the current design, which limited measurement time lags to a 1-year unit. Future studies might well employ time units that are theoretically selected to represent the unit of hypothesized change in a developmental sequence, even allowing for unequal time units if theory dictates.

Conclusions

Even with these numerous caveats, the current study offers a novel dynamic cascade model of the development of serious violent behavior in adolescence, innovative use of methods of aggregating and analyzing developmental data, and empirical findings that inform developmental theory as well as preventive intervention. These methods await application to other developmental phenomena.

References

Aber, J. L., Gershoff, E., Ware, A., & Kotler, J. (2004). Estimating the effects of September 11th, 2001, and other forms of violence on the mental health and social development of New York City's youth: A matter of context. *Applied Developmental Science, 8*, 111–129.

Achenbach, T. M. (1991). *Manual for the Child Behavior Checklist/4–18 and 1991 profile.* Burlington: University of Vermont Department of Psychiatry.

Beyers, J. M., Bates, J. E., Pettit, G. S., & Dodge, K. A. (2003). Neighborhood structure, parenting processes, and the development of youths' externalizing behaviors: A multilevel analysis. *American Journal of Community Psychology, 31*, 35–53.

Capaldi, D. M., & Patterson, G. R. (1996). Can violent offenders be distinguished from frequent offenders: Prediction from childhood to adolescence. *Journal of Research in Crime and Delinquency, 33*, 206–231.

Caspi, A., Moffitt, T. E., Morgan, J., Rutter, M., Taylor, A., Arseneault, L., et al. (2004). Maternal expressed emotion predicts children's antisocial behavior: Using MZ-twin differences to identify environmental effects on behavioral development. *Developmental Psychology, 40*, 149–161.

Chin, W. W. (1995). Partial least squares is to LISREL as principal components analysis is to common factor analysis. *Technology Studies*, 2, 315–319.

Coie, J. D., Dodge, K. A., & Coppotelli, H. (1982). Dimensions and types of social status: A cross-age perspective. *Developmental Psychology, 18*, 557–570.

Conduct Problems Prevention Research Group. (1992). A developmental and clinical model for the prevention of conduct disorders: The FAST Track Program. *Development and Psychopathology, 4*, 509–527.

Conduct Problems Prevention Research Group. (1999). Initial impact of the Fast Track Prevention Trial for Conduct Problems: I. The high-risk sample. *Journal of Consulting and Clinical Psychology, 67*, 631–647.

Conduct Problems Prevention Research Group. (2004). The effects of the Fast Track Program on serious problem outcomes at the end of elementary school. *Journal of Clinical Child and Adolescent Psychology, 33*, 650–661.

Crnic, K., & Greenberg, M. (1990). *Interaction Ratings Scale.* Unpublished technical report, Pennsylvania State University.

Cummings, E. M., & Davies, P. T. (2002). Effects of marital conflict on children: Recent advances and emerging themes in process-oriented research. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 43*, 31–63.

Dishion, T. J., & Kavanagh, K. (2000). A multilevel approach to family-centered prevention in schools: Process and outcome. *Addictive Behaviors, 25*, 899–911.

Dishion, T. J., & McMahon, R. J. (1998). Parental monitoring and the prevention of child and adolescent problem behavior: A conceptual and empirical formulation. *Clinical Child and Family Psychology Review, 1*, 61–75.

Dishion, T. J., Nelson, S. E., & Bullock, B. M. (2004). Premature adolescent autonomy: Parent disengagement and deviant peer process in the amplification of problem behavior. *Journal of Adolescence, 27*, 515–530.

Dodge, K. A., Bates, J. E., & Pettit, G. S. (1990). Mechanisms in the cycle of violence. *Science, 250*, 1678–1683.

Dodge, K. A., Coie, J. D., & Lynam, D. (2006). Aggression and antisocial behavior in youth. In W. Damon (Series Ed.) & N. Eisenberg (Vol Ed.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (6th ed., pp. 719–788). New York: Wiley.

Dodge, K. A., Lansford, J. E., Burks, V. S., Bates, J. E., Pettit, G. S., Fontaine, R., et al. (2003). Peer rejection and social information-processing factors in the development of aggressive behavior problems in children. *Child Development, 74*, 374–393.

Dodge, K. A., Malone, P. S., Lansford, J. E., Miller-Johnson, S., Pettit, G. S., & Bates, J. E. (2006). Toward a dynamic developmental model of the role of parents and peers in early onset substance use. In A. Clarke-Stewart & J. Dunn (Eds.), *Families count: Effects on child and adolescent development* (pp. 104–134). New York: Cambridge University Press.

Dodge, K. A., & Pettit, G. S. (2003). A biopsychosocial model of the development of chronic conduct problems in adolescence. *Developmental Psychology, 39*, 349–371.

Dodge, K. A., Pettit, G. S., & Bates, J. E. (1994). Socialization mediators of the relation between socioeconomic status and child conduct problems. *Child Development, 65*, 649–665.

Dodge, K. A., Pettit, G. S., McClaskey, C. L., & Brown, M. (1986). Social competence in children. *Monographs of the Society for Research in Child Development, 51*(2, Serial No. 213).

Eisenberg, N., Spinrad, T. L., Fabes, R. A., Reiser, M., Cumberland, A., Shepard, S. A., et al. (2004). The relations of effortful control and impulsivity to children's resiliency and adjustment. *Child Development, 75*, 1–22.

Eyberg, S. M. (1974). *Manual for the coding of parent-child interactions of children with behavior problems*. Unpublished manuscript, Oregon Health Sciences University.

Glueck, S., & Glueck, E. (1950). *Unraveling juvenile delinquency*. Cambridge, MA: Harvard University Press.

Gorman-Smith, D., & Loeber, R. (2005). Are developmental pathways in disruptive behaviors the same for girls and boys? *Journal of Child and Family Studies, 14*, 15–27.

Gottfredson, M. R., & Hirschi, T. (1990). *A general theory of crime*. Stanford, CA: Stanford University Press.

Graham, J. W., Hofer, S. M., & MacKinnon, D. P. (1996). Maximizing the usefulness of data obtained with planned missing value patterns: An application of maximum likelihood procedures. *Multivariate Behavioral Research, 31*, 197–218.

Granic, I., & Patterson, G. R. (2006). Toward a comprehensive model of antisocial development: A dynamic systems approach. *Psychological Review, 113*, 101–131.

Greenberg, M. L., Lengua, L. S., Coie, J., Pinderhughes, E. E., & the Conduct Problems Prevention Research Group. (1999). Predicting developmental outcomes at school entry using a multiple-risk model: Four American communities. *Developmental Psychology, 35*, 403–417.

Henggeler, S. W., Schoenwald, S. K., Bourduin, C. M., Rowland, M. D., & Cunningham, P. B. (1998). *Multisystemic treatment of antisocial behavior in children and adolescents*. New York: Guilford.

Hill, N. E., Castellino, D. R., Lansford, J. E., Nowlin, P., Dodge, K. A., Bates, J. E., et al. (2004). Parent-academic involvement as related to school behavior, achievement, and aspirations: Demographic variations across adolescence. *Child Development, 75*, 1491–1509.

Hinshaw, S. P. (1992). Externalizing behavior problems and academic underachievement in childhood and adolescence: Causal relationships and underlying mechanisms. *Psychological Bulletin, 111*, 127–155.

Hollingshead, A. B. (1975). *Four factor index of social status*. Unpublished manuscript, Yale University, New Haven, CT.

Holmes, C. T. (1989). Grade level retention effects: A meta-analysis of research studies. In L. A. Shepard & M. L. Smith (Eds.), *Flunking grades: Research and policies on retention* (pp. 16–33). Philadelphia: Falmer Press.

Huizinga, D., & Elliott, D. S. (1987). Juvenile offenders: Prevalence, offender incidence, and arrest rates by race. *Crime and Delinquency, 33*, 206–223.

Ingoldsby, E. M., & Shaw, D. S. (2002). Neighborhood contextual factors and early-starting antisocial pathways. *Clinical Child and Family Psychology Review, 5*, 21–55.

Keenan, K., Loeber, R., Zhang, Q., Stouthamer-Loeber, M., & Van Kammen, W. B. (1995). The influence of deviant peers on the development of boys' disruptive and delinquent behavior: A temporal analysis. *Development and Psychopathology, 7*, 715–726.

Laird, R. D., Jordan, K. Y., Dodge, K. A., Pettit, G. S., & Bates, J. E. (2001). Peer rejection in childhood, involvement with antisocial peers in early adolescence, and the development of externalizing behavior problems. *Development and Psychopathology, 13*, 337–354.

Laird, R. D., Pettit, G. S., Bates, J. E., & Dodge, K. A. (2003). Parents' monitoring-relevant knowledge and adolescents' delinquent behavior: Evidence of correlated developmental changes and reciprocal influences. *Child Development, 74*, 752–768.

Lansford, J. E., Deater-Deckard, K., Dodge, K. A., Bates, J. E., & Pettit, G. S. (2004). Ethnic differences in the link between physical discipline and later adolescent externalizing behaviors. *Journal of Child Psychology and Psychiatry, 45*, 801–812.

Lansford, J. E., Dodge, K. A., Pettit, G. S., Bates, J. E., Crozier, J., & Kaplow, J. (2002). A 12-year prospective study of the long-term effects of early child physical maltreatment on psychological, behavioral, and academic problems in adolescence. *Archives of Pediatrics and Adolescent Medicine, 156*, 824–830.

Larzelere, R. E., & Patterson, G. R. (1990). Parental management: Mediators of the effect of socioeconomic status on early delinquency. *Criminology, 28*, 301–323.

Lengua, L. J., Wolchik, S. A., Sandler, I. N., & West, S. G. (2000). The additive and interactive effects of temperament and parenting in predicting adjustment problems of children of divorce. *Journal of Clinical Child Psychology, 29*, 232–244.

Lochman, J. E., & the Conduct Problems Prevention Research Group. (1995). Screening of child behavior problems for prevention programs at school entry. *Journal of Consulting and Clinical Psychology, 63*, 549–559.

Loeber, R., Farrington, D. P., Stouthamer-Loeber, M., & van Kammen, W. B. (1998). *Antisocial behavior and mental health problems: Explanatory factors in childhood and adolescence*. Mahwah, NJ: Erlbaum.

Loeber, R., Pardini, D., Hornish, D. L., Wei, E. H., Crawford, A. M., Farrington, D. P., et al. (2005). The prediction of violence and homicide in young men. *Journal of Consulting and Clinical Psychology, 73*, 1074–1088.

Lohmoller J.-B. (1989): *Latent variables path modeling with partial least squares.* Heidelberg, Germany: Physica-Verlag.

Lynam, D. R., Caspi, A., Moffitt, T. E., Wikström, P. O., Loeber, R., & Novak, S. P. (2000). The interaction between impulsivity and neighborhood context on offending: The effects of impulsivity are stronger in poorer neighborhoods. *Journal of Abnormal Psychology, 109*, 563–574.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test the significance of the mediated effect. *Psychological Methods, 7*(1), 83–104.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2006). Frequency distribution, 10,000 replications, no mediated effect, all continuous variables [on-line]. Retrieved November 22, 2006, from http://www.public.asu.edu/~davidpm/ripl/freqdist.pdf

McLoyd, V. (1990). The impact of economic hardship on black families and children: Psychological distress, parenting, and socioemotional development. *Child Development, 61*, 311–346.

Moffitt, T. E. (1993). Adolescence-limited and life-course-persistent antisocial behavior: A development taxonomy. *Psychological Review, 100*, 674–701.

Moffitt, T. E., & Caspi, A. (2001). Childhood predictors differentiate life-course persistent and adolescence-limited antisocial pathways among males and females. *Development and Psychopathology, 13*, 135–151.

Moffitt, T. E., Gabrielli, W. F., Mednick, S. A., & Schulsinger, F. (1981). Socioeconomic status, IQ, and delinquency. *Journal of Abnormal Psychology, 90*, 152–156.

Muthén, L. K., & Muthén, B. O. (2006). *Mplus user's guide, v.4*. Los Angeles, CA: Author.

O'Donnell, J., Hawkins, J. D., & Abbott, R. D. (1995). Predicting serious delinquency and substance use among aggressive boys. *Journal of Consulting and Clinical Psychology, 63*, 529–537.

Patterson, G. R. (1986). Performance models for antisocial boys. *American Psychologist, 41*, 432–444.

Patterson, G. R., Reid, J. B., & Dishion, T. J. (1992). *A social learning approach: Vol. 4. Antisocial boys*. Eugene, OR: Castalia Press.

Pettit, G. S., Bates, J. E., & Dodge, K. A. (1997). Supportive parenting, ecological context, and children's adjustment. *Child Development, 68*, 908–923.

Pettit, G. S., Bates, J. E., Dodge, K. A., & Meece, D. W. (1999). The impact of after-school peer contact on early adolescent externalizing problems is moderated by parental monitoring, neighborhood safety, and prior adjustment. *Child Development, 70*, 768–778.

Pinderhughes, E. E., Nix, R., Foster, E. M., Jones, D., & the Conduct Problems Prevention Research Group. (2001). Parenting in context: Impact of neighborhood poverty, residential stability, public services, social networks, and danger on parental behaviors. *Journal of Marriage and the Family, 63*, 941–953.

Pollak, S. D., & Tolley-Schell, S. A. (2003). Selective attention to facial emotion in physically abused children. *Journal of Abnormal Psychology, 112*, 323–338.

Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*, 385–401.

Ribordy, S. C., Camras, L. A., Stefani, R., & Spaccarelli, S. (1988). Vignettes for emotion recognition research and affective therapy with children. *Journal of Clinical Child Psychology, 17*, 322–325.

Rutter, M., & Garmezy, N. (1983). Developmental psychopathology. In P. H. Mussen & E. M. Hetherington (Eds.), *Handbook of child psychology: Vol. 4. Socialization, personality and social development* (pp. 775–911). New York: Wiley.

Sameroff, A. J., Bartko, W. T., Baldwin, A., & Baldwin, C. (1988). *Families, risk, and competence*. Mahwah, NJ: Erlbaum.

Sameroff, A. J., & Mackenzie, J. (2003). Research strategies for capturing transactional models of development: The limits of the possible. *Development and Psychopathology, 15*, 613–640.

Sampson, R. J., & Laub, J. H. (1994). Urban poverty and the family context of delinquency: A new look at structure and process in a classic study. *Child Development, 65*, 523–540.

Sampson, R. J., Raudenbush, S. W., & Earls, F. (1997). Neighborhoods and violent crime: A multi-level study of collective efficacy. *Science, 277*, 918–924.

SAS Institute. (2004). *SAS user's manual, v.9.1.3*. Cary, NC: Author.

Simons, R. L., Wu, C. I., Conger, R. D., & Lorenz, F. O. (1994). Two routes to delinquency: Differences between early and late starters in the impact of parenting and deviant peers. *Criminology, 32*, 247–276.

Snyder, J., Dishion, T. J., & Patterson, G. R. (1986). Determinants and consequences of associating with deviant peers during preadolescence and adolescence. *Journal of Early Adolescence, 6*, 29–43.

Snyder, J., Reid, J., & Patterson, G. (2003). A social learning model of child and adolescent antisocial behavior. In B. B. Lahey, T. E., Moffitt, & A. Caspi (Eds.), *Causes of conduct disorder and juvenile delinquency* (pp. 27–48). New York: Guilford.

Stattin, H., & Kerr, M. (2000). Parental monitoring: A reinterpretation. *Child Development*, *71*, 1072–1085.

Stouthamer-Loeber, M., Loeber, R., Wei, E., Farrington, D., & Wikström, P. H. (2002). Risk and promotive effects in the explanation of persistent serious delinquency in boys. *Journal of Consulting and Clinical Psychology*, *70*, 111–123.

Straus, M. A. (1979). Measuring intrafamily conflict and violence: The Conflict Tactics Scale. *Journal of Marriage and the Family*, *41*, 75–88.

Strayhorn, J. M., & Weidman, C. S. (1988). A parenting practices scale and its relation to parent and child mental health. *Journal of the American Academy of Child and Adolescent Psychiatry*, *27*, 613–618.

Tremblay, R. E., Pagani-Kurtz, L., Masse, L. C., Vitaro, F., & Pihl, R. O. (1995). A bi-modal preventive intervention for disruptive kindergarten boys: Its impact through mid-adolescence. *Journal of Consulting and Clinical Psychology*, *63*, 560–568.

Wold, H. (1982). Soft modeling: The basic design and some extensions. In K. G. Joreskog & H. Wold (Eds.), *Systems under indirect observation: Causality, structure, prediction* (Vol. 2, pp. 1–54). Amsterdam: North-Holland.

Child Development, November/December 2008, Volume 79, Number 6, Pages 1928–1941

Investigating Preschoolers' Categorical Thinking About Gender Through Imitation, Attention, and the Use of Self-Categories

Diana M. Grace and Barbara J. David
The Australian National University

Michelle K. Ryan
The University of Exeter

Whereas traditional theories of gender development have focused on individualistic paths, recent analyses have argued for a more social categorical approach to children's understanding of gender. Using a modeling paradigm based on K. Bussey and A. Bandura (1984), 3 experiments ($N = 62$, $N = 32$, and $N = 64$) examined preschoolers' (M age = 52.9 months) imitation of, and memory for, behaviors of same-sex and opposite-sex children and adults. In all experiments, children's imitation of models varied according to the emphasis given to the particular category of models, despite equal attention being paid to both categories. It is suggested that the categorical nature of gender, or age, informs children's choice of imitative behaviors.

Few things characterize children's lives as much as gender. From birth, children's environments and experiences differ based purely on their sex (Condry & Condry, 1976; Rheingold & Cook, 1975; Rubin, Provenzano, & Luria, 1974). Boys and girls receive differential treatment both at home (Siegal, 1987) and at school (Tittle, 1986), and this occurs against a backdrop of media portrayals further differentiating between the roles and status of men and women (Durkin, 1985).

Children themselves also play an active role in perpetuating this ubiquitous role of gender. At around 30–36 months of age, children display a marked preference for same-sex peers (LaFreniere, Strayer, & Gauthier, 1984; Maccoby & Jacklin, 1987; Powlishta, Serbin, & Moller, 1993), with these preferences increasing throughout childhood. By preschool age, 50% of children's play involves only opposite-sex peers (Fabes, Martin, & Hanish, 2003; Maccoby & Jacklin, 1987). Such same-sex preferences occur independent of parental, or other adult, interactions; they increase when adults are not present (Maccoby & Jacklin, 1987; Thorne, 1986); and they appear resistant to attempts to change (Serbin, Tonick, & Sternglanz, 1977). Although girls tend to initiate same-sex play more than boys at age 3 (Serbin, Moller, Gulko, Powlishta, & Colburne, 1994), this gender difference is reversed, and even heightened, by age 5 (Pitcher & Schultz, 1983).

The phenomenon of same-sex preference also becomes one of cross-sex avoidance, as it becomes equally important not to play with members of the opposite sex as it is to play with members of the same sex (e.g., Feiring & Lewis, 1987). This is particularly the case for boys who, even in the toddler years, actively avoid "not male" activities (Fagot, 1985a). This proclivity for gender differentiation often leads to a view of gender differences as inexorable and inevitable (a view that is explicitly challenged by the current research) and has played a role in theoretical explanations looking to within-sex rather than between-sex differences.

Despite the recognition that sex differences, and the display of sex-typed preferences, tend to occur in group settings, theories devoted to understanding gender development traditionally focused on gender development as a uniquely individual process (e.g., Bem, 1981, 1984; Freud, 1924, 1933; Kohlberg, 1966; Martin & Halverson, 1981, 1983). Being male or female, however, is not only about the individual; it implies membership of one category or another. Such explicit recognition of the group, or categorical, nature of gender has influenced more recent theoretical explanations and helps to explain why the adoption of gender norms and behaviors is essentially unrelated to individual attributes, or degree of sex typing, of the children who comprise these groups (Maccoby & Jacklin, 1987; Powlishta, 1989, cited in Maccoby, 1990). Moreover, the different experiences that boys and girls have in their same-sex groups may actually reinforce the importance of gender as an organizing property of these peer groups (Fabes et al., 2003; Maccoby, 1990) and may supersede innate differences or preference for playstyles that may have prompted the sex segregation (Fabes et al., 2003; Martin & Fabes, 2001).

Correspondence concerning this article should be addressed to Diana M. Grace, Department of Psychology, The Australian National University, Canberra ACT 0200, Australia. Electronic mail may be sent to diana.grace@anu.edu.au.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7906-0026

In addition to organizing information, gender can be used to inform children about appropriate behaviors—an idea that may help explain why young children imitate others of the same sex as themselves. This seems to be particularly the case for boys who appear just as intent on not imitating girls as they are on imitating boys and are more likely than girls to reject, or tease, each other for engaging in opposite-sex behavior (Feiring & Lewis, 1987). Again, however, such a view assumes that gender effects operate in a fixed way. The current research was designed to explore this idea by examining the notion of similarity to peers (and/or same-sex adults) and the contexts in which (gender-relevant) information is provided. The social categorical, or group-based, nature of gender is emphasized throughout.

Central to this idea of gender as a group phenomenon is work by Bussey and colleagues (Bussey & Bandura, 1984; Bussey & Perry, 1982) whose modeling studies demonstrate that adoption of same-sex behaviors, and rejection of opposite-sex ones, depended on those behaviors being seen as normative, or not, for a number of men or women: It appeared that children waited to see if a particular behavior was exhibited by all the men or all the women before adopting (or rejecting) it for themselves. This, the authors argue, reflects how children see their social environment. On any given day, children are exposed to numerous people exhibiting a variety of behaviors. What they encode as gender normative, however, are those behaviors exhibited most frequently and consistently by multiple members of each gender category.

This is similar to a more extreme idea put forward by Harris (1995, 1998) who claims that children are influenced more by other children than by parents or other adults. She emphasizes the group basis of lasting influence by making a clear distinction between dyadic friendships and group memberships (Harris, 1998). The effects of dyadic relationships are claimed to be short-lived, with any long-term influence attributed to the fact that friends tend to be members of the same social groups. Thus, Harris (1998) claims it is the processes functioning within the child's peer group that are responsible for the lasting transmission of culture and the resultant adult outcomes.

This approach, however, relies on a similarity-based account of identification. That is, because children see themselves as similar to other children, or because girls (or boys) see themselves as similar to other girls (or boys), they identify with, and are influenced by, these similar others or those who share category membership. Such an analysis, however, views both similarity and difference as fixed features of a stimulus.

Gender as a Group Process—The Role of Categorization

The role of categorization has been increasingly implicated in empirical investigations concerning young children's social category knowledge, although these have typically focused on primary school–aged children (e.g., Bigler, Jones, & Lobliner, 1997). These investigations have included such domains as gender (e.g., Powlishta, 1995b), race and ethnicity (e.g., Nesdale & Flesser, 2001; Verkuyten, 2002; Yee & Brown, 1992), national identity (e.g., Barrett, 2000), and, to some degree, essentialist beliefs about categories (e.g., Heyman & Gelman, 2000). Much of this work has been explicitly informed by social identity theory (SIT; Tajfel & Turner, 1986) and, to a lesser extent, self-categorization theory (SCT; Turner, Hogg, Oakes, Reicher, & Wetherell, 1987).

According to this approach, social categorization is both adaptive—in that it provides information about the social world—and a reflection of our active engagement with that social world (e.g., Oakes, Turner, & Haslam, 1991). This view thus provides a crucial link between external differences (e.g., gender differences) and how these can be reflected, and represented, within the person.

An important outcome of the categorization process that has received much attention is the perceptual distortion (such as the accentuation of similarity and difference) that pertains to (same-sex) similarity or (same-sex) preference. Although accentuation effects have been demonstrated using racial categories (Hewstone, Hantzi, & Johnston, 1991; Taylor, Fiske, Etcoff, & Ruderman, 1978), gender categories (Frable & Bem, 1985; Jackson & Hymes, 1985), and even sexual preference (Walker & Antaki, 1986), these effects need not necessarily be considered an automatic outcome of the categorization process but can vary according to the appropriateness and meaningfulness of the salient categorization (McGarty & Penny, 1988; McGarty & Turner, 1992).

These principles of accentuation (of similarities and differences) have also been applied to categorizations of the self (Turner, 1984), whereby self-categorization in terms of a social identity maximizes both differences between one's in-group and an out-group and similarities between oneself and other in-group members. Similarly, self-categorization in terms of one's personal identity maximizes perceived differences between oneself and other individuals while simultaneously minimizing intraindividual (within oneself) inconsistencies.

Such an approach explicitly rejects the notion of similarity (or difference) as a fixed feature of a stimulus (cf. Medin, Goldstone, & Gentner, 1993; Murphy &

Medin, 1985; Yzerbyt, Rocher, & Schadron, 1997). Similarities (within categories) and differences (between categories) are guided not only by a comparative frame of reference but also by the perceiver's own expectations and theories (Brown & Turner, 2002). Children categorize children and adults (and boys and girls) not only because of perceptual similarities and differences but because there is some psychological reason to do so. They expect children and adults (and boys and girls) to look and behave differently partly because they have seen them do so and partly because they have seen the consequences of these different behaviors.

Children's Understanding of Social Categories

Applying a self-categorization approach to children's understanding of themselves and others suggests that children define themselves in a number of ways. They may regard themselves as children, as boys (or girls), as members of their family (the Johnsons or the Billington-Smythes), or even as "tadpoles" (in their learn to swim class). This self-definition is determined by both the current situation and the children's own expectations about these categorizations.

Such an approach has indeed been recently articulated by Bigler and Liben (2006, 2007) whose developmental intergroup theory (DIT) outlines the mechanisms by which children develop social stereotypes and prejudiced beliefs. This work is consistent with the adult social psychological literature in focusing on the social psychological processes implicated in such thoughts and behaviors and in emphasizing the generality of these processes. That is, these psychological processes are not confined to the domain of racial, or ethnic, prejudice but apply to the range of intergroup, or intercategory, domains, including gender.

The Current Research

The overall aim of the current research, therefore, was an empirical investigation of social categorization (specifically, the notions of variable self-definition, similarity as a reflection of perceptual and contextual features, and the effect of salience manipulations) in the domain of gender development. The three studies presented focused explicitly on examining the acquisition of gender understanding as a social category. We sought to investigate how this particular social category functions, whether it differs from other social categories (in particular, age), and what the implications are of belonging to one of those social categories (e.g., male as opposed to female). Given that children are capable of using one of a number of social categories, we suggest that the importance of gender to children may be a consequence of the importance it is afforded, that is, its relative salience within a given context.

Relevant processes in this investigation included the imitation of observed others and the attention paid to information presented. Given its importance in the socialization process, modeling, or imitation, was a major focus of this work using a modeling paradigm (after Bussey & Bandura, 1984), which we describe subsequently.

Experiment 1

The first experiment aimed to extend Bussey and Bandura's (1984) findings of sex-linked modeling behaviors. In this earlier work, children watched videos in which a different set of behaviors was performed by adult male and female models before being given an opportunity to enact these behaviors. Although no social context had been provided for the behaviors, children showed a strong preference for imitating the same-sex models, despite demonstrating equal recall of the behaviors performed by both the male and the female models. Given the increased emphasis since the original study on nongendered play and education in early childhood settings (e.g., Yellond, 1998) and the expression of this by early childhood teachers (e.g., Cahill & Adams, 1997), the first aim was to investigate whether these preferences still exist.

The second, and related, aim was to investigate children's sex-linked modeling with peers, or other children, not only to determine the robustness of sex-linked modeling but also to pose a direct test of Harris's (1995, 1998) assertions that children are more influential models than adults. Although recognizing that this was in no way an attempt to measure lasting influence, the comparison allowed us to examine variable self-categorization in children as a result of changes in contextual features.

A measure of attention paid to models provided important information for children's imitative behaviors. If, for example, children imitate same-sex models more than opposite-sex ones, this does not tell us whether children simply ignored the information provided by the opposite sex in the first instance or if they paid attention to all the information provided. The former suggests a relatively straightforward process involving identifying self and others (models) as male or female and then finding it appropriate to imitate the behavior of the person similar to themselves (the same-category exemplar). If, however, the latter is the case, it implies a more complex process involving consciously choosing an appropriate behavior from a wider stimulus array. It also implies that some behaviors are consciously rejected as being inappropriate behaviors to enact.

In line with Bussey and Bandura (1984), it was predicted that both boys and girls would imitate same-sex models more than opposite-sex models. Although Bussey and Bandura found greater same-sex modeling by boys, this would not be expected if boys and girls use gender categories in the same way.

With respect to modeling of adults and children, an extrapolation of Harris' (1998) argument would predict more modeling of child models than of adult models, although more contemporary social developmental understandings (e.g., DIT, extrapolations of SCT) would predict no difference between these (at least in situations of low gender salience and/or low comparative context). Because these predictions have not previously been investigated, this was of particular interest in the current study.

Finally, age of participants was analyzed to investigate any developmental change that might occur, even within the limited age span. It was anticipated that older children would have had more exposure to gender socialization pressures and this would manifest itself in greater demonstration of same-sex modeling than by younger children.

Method

Participants. Participants were 32 girls and 30 boys from university child-care centers in Canberra, Australia. They ranged in age from 39 to 68 months ($M = 53.98$ months, $SD = 6.48$). No specific data on children's socioeconomic status or race/ethnicity were obtained, although the majority of children attending these centers have parents either working or studying at the university and represent a somewhat diverse range of ethnic backgrounds. Participants had written parental consent and provided verbal consent themselves at the start of each session. All participants were able to understand and speak English.

Design. The study involved a 2 (sex of participant: boy, girl) × 2 (age of models: adult, child) × 2 (sex of models: male, female) mixed factorial design. The last factor was the within-participants factor, as all children saw videos of both male and female models.

Procedure. Participants watched one of four videos depicting three male and three female models (either all adults or all children), where all the males performed one set of behaviors and all the females another. The behaviors performed by the male and female actors were counterbalanced in alternative versions of the videos. As with the child participants, the models themselves came from a range of ethnic backgrounds, although this was not emphasized in any way, nor was it commented on by any child.

After being introduced to the experimenter in a classroom situation, each participant was individually brought into the experimental room (a separate room at each child-care center). The room was set up in a manner identical to what they would see in the video, with chairs, a table with two Mickey Mouse hats, another table with a box and a toy koala, a pinboard on one wall, and a television. The child sat next to the experimenter to watch the video.

The sequences of behaviors performed by the models were those described in Bussey and Bandura (1984). In one sequence, all the male actors chose a black Mickey Mouse hat, which they placed on their head with the Mickey Mouse symbol to the front and performed a corresponding behavior sequence. In the alternative sequence, actors chose a red Mickey Mouse hat that they wore with the Mickey Mouse symbol to the back and performed the alternative behavior sequence. Each sequence of behaviors comprised 14 separate behavioral components.

After watching the entire video, each child was asked by the experimenter if he or she would like to play the game he or she had just seen. The experimenter reminded each child first to choose a hat, the only prompt provided at this stage of the study. Each child was then allowed to play with the toys as shown in the video. This behavior served as the primary "imitation" dependent measure. Where parental consent had been provided, the child's performance was videotaped.

Memory task/manipulation check. Following the imitation task, each child was asked to recall the behaviors of the male and female models. For each of the 14 behaviors, the experimenter provided participants with specific prompts, such as "Which hat did the men (boys)/women (girls) choose?" The order of questions (male followed by female or female followed by male) was counterbalanced across participants. Responses were recorded on a checklist by the experimenter.

At the end of each session, the children were thanked for their participation and allowed to choose a sticker to take home.

Results

Modeled (imitation) behavior. A checklist of the behaviors was used to record each child's performance. The total number of behaviors was calculated for each of the four categories: (a) same-sex behavior, (b) opposite-sex behavior, (c) other (i.e., new or random) behavior, and (d) nothing (i.e., no behavior performed). The final category was not used in the analyses because it was of no theoretical interest and provided no additional information to what was provided by the other categories.

Videos of 39 participants were analyzed by an independent rater using the same four categories. Using Cronbach's alpha, reliability coefficients for each of the 14 behaviors ranged from .71 to 1.00. In an attempt to reach perfect agreement, both raters reanalyzed the videos. Disagreements occurred only when tape quality was poor. The subsequent studies were not videotaped due to issues concerning parental consent and the sheer impracticality of the procedure. Given the high accuracy of the primary rater, however, this was not considered a problem.

The data were analyzed in a 2 (sex of participant: boy, girl) × 2 (age of model: adult, child) × 3 (modeled behaviors: same sex, opposite sex, new behavior) analysis of variance (ANOVA), with the last factor being within participants. Means and standard deviations appear in Table 1. A significant main effect for modeled behaviors, $F(2, 116) = 20.38, p < .001$, partial $\eta^2 = .26$, indicated greater performance of same-sex modeling ($M = 3.31$) than either opposite-sex modeling ($M = 1.58$) or performance of a new behavior ($M = 1.28$). This was qualified by a significant Modeled Behaviors × Age of Model interaction, $F(2, 116) = 3.04, p = .05$, partial $\eta^2 = .05$. Children imitated same-sex behaviors more when they were modeled by adults ($M = 3.89$) than when they were modeled by children ($M = 2.73$), $t(60) = 1.95, p = .05$; children responded (nearly) equally to opposite-sex adult ($M = 1.58$) and child ($M = 1.59$) models and performed new behaviors more when models were children ($M = 1.52$) than when they were adults ($M = 1.04$), although this difference did not reach significance, $t(60) = -1.83, p = .07$. The Modeled Behaviors × Sex of Participant interaction was nonsignificant, $F(2, 116) = 2.30, p = .11$, partial $\eta^2 = .04$.

Table 1

Experiment 1: Mean Number (Standard Deviations) of Modeled Behaviors of Adult and Child Models

	Adult models		Child models	
	Girls	Boys	Girls	Boys
Same sex	4.38 (2.90)	3.40 (2.29)	3.19 (2.01)	2.27 (2.02)
Opposite sex	1.63 (1.54)	1.53 (1.81)	1.38 (1.54)	1.80 (1.66)
New behavior	0.75 (0.93)	1.53 (0.98)	1.38 (1.26)	1.67 (0.90)

Although the overall three-way interaction was statistically nonsignificant, our precise hypotheses led us to explore separately imitation of adult and child models.

Imitation of adult models. Both girls and boys imitated same-sex adult models significantly more than opposite-sex adult models, $F(2, 58) = 16.75, p < .001$, partial $\eta^2 = .37$. However, the linear decrease in behaviors from same-sex imitation through opposite-sex imitation to new behaviors was evident only for girls. Whereas girls' performance of opposite-sex behaviors was significantly greater than their performance of new behaviors, $t(15) = 2.05, p = .05$, boys were no more likely to perform opposite-sex behaviors than new behaviors, $t(14) = 0.34, p = .74$. In other words, boys were no more likely to imitate a woman than to invent entirely new behaviors.

Imitation of child models. The pattern of imitating child models also differed for girls and boys, $F(2, 58) = 4.69, p = .013$, partial $\eta^2 = .14$. Although girls' same-sex modeling was significantly greater than their opposite-sex modeling, $t(15) = 2.72, p = .01$, their performance of opposite-sex modeling was identical to two decimal places to that of new behaviors. Strikingly, however, the amount of behaviors in each of the three categories did not significantly differ for boys, $F(2, 28) = 0.52, p = .60$. Boys were no more likely to imitate same-sex child models than they were to imitate opposite-sex child models or to perform a completely new behavior.

Age of participants. To investigate the possibility of a developmental increase in imitation of models, even at this young age, the above analyses were again performed to include age of participants. A median split was used to divide older from younger participants: the median age was 55 months. Again, the strong main effect for modeled behaviors emerged, $F(2, 108) = 15.39, p < .001$, partial $\eta^2 = .22$. Age of participant did not enter into any significant main or interaction effects.

In summary, there was clear evidence of children's same-sex imitation that, contrary to Bussey and Bandura (1984), did not statistically differ for girls and boys. Importantly for the present study and for Harris's (1995, 1998) contention concerning the prime importance of peers, children imitated behaviors performed by adult models more than when those same behaviors were performed by child models.

Memory task/manipulation check. For this task, children were specifically asked what actions the male or female models had performed. The data were summed to produce the following: (a) correct recall of same-sex models' behavior, (b) correct recall of opposite-sex models' behavior, (c) incorrect recall of same-sex models' behavior, and (d) incorrect recall of opposite-sex models' behavior. The data were analyzed in a 2 (sex of participant: boy, girl) × 2 (age of model: adult, child) × 4 (memory for models) mixed-model ANOVA, with memory for modeled behavior as a within-participants factor. Means and standard deviations appear in Table 2.

Strong main effects for memory for modeled behavior, $F(3, 174) = 21.28, p < .001$, partial $\eta^2 = .27$, and for age of model, $F(1, 58) = 15.58, p < .001$, partial $\eta^2 = .21$, were qualified by a significant Memory for Modeled Behavior × Age of Model interaction, $F(3, 174) = 35.11, p < .001$, partial $\eta^2 = .12$. Of particular interest, children's correct recall of same-sex adult models' behavior ($M = 6.13$) was superior to their correct recall of those same behaviors when performed by child models ($M = 3.65$), $t(60) = 4.28, p < .001$. Similarly, their correct recall of opposite-sex models' behaviors was superior for adult models ($M = 5.75$) than for child models ($M = 3.55$), $t(60) = 3.80, p < .001$. Children's incorrect memory did not differ for adult and child models. Rather than being affected by the sex of model, children's recall of behaviors, or more specifically their correct recall of behaviors, was affected by the age of the models enacting those behaviors, with more correct recall of adult behaviors. No other main or interaction effects were found. As with the imitation measures, the analyses were repeated including age of participants. No effects were found.

Discussion

Consistent with Bussey and Bandura (1984), children showed a strong preference for imitating same-sex models. The present study extended the previous work by showing that this preference for same-sex imitation was maintained when models were children as well as when they were adults. It must be noted, however, that the frequency of behaviors not performed was quite high (over half of all possible behaviors); we address this overall rate of imitation in our next studies.

Table 2
Experiment 1: Mean Number (Standard Deviations) of Recalled Behaviors of Adult and Child Models

	Adult models	Child models
Same sex correct	6.13 (2.55)	3.65 (1.98)
Opposite sex correct	5.75 (2.59)	3.55 (1.89)
Same sex incorrect	2.35 (1.11)	2.74 (1.14)
Opposite sex incorrect	2.67 (1.58)	2.78 (1.03)

Bussey and Bandura's (1984) finding of greater same-sex modeling by boys than by girls was not replicated in the current study. Furthermore, and contrary to predictions derived from Harris (1995, 1998), children imitated same-sex models more when those models were adults than when they were children.

Clearly, however, children responded to the gender categories provided. Specifically, we argue that children chose to enact the behaviors performed by the category of people they saw as most appropriate to imitate. With respect to gender, it could also be said that the children saw same-sex people as most similar to themselves. However, with respect to age categories it could not be argued that the children saw themselves as more similar to the other children and hence decided to imitate them to a greater extent. Indeed, the current findings indicate the reverse: Children imitated the adult models to a greater extent than the child models, suggesting that (perceptual) similarity to other children was not sufficient to produce imitation of perceptually similar (i.e., same-age) models. This is inconsistent with predictions by Harris (1995, 1998).

It is possible, however, that the children recognized the authority of adults and hence saw it appropriate to imitate adults' actions. Children also have considerable experience imitating adults, especially with respect to gender-related activities. However, given the amount, and intensity, of reinforcement provided by peers for gender stereotypical behavior—even at the age of 3 or 4 years (e.g., Huston, 1983), along with the amount of time children spend in unstructured freeplay sessions, it is unlikely that they would have more experience imitating adults than children.

What makes the current findings particularly remarkable is that the behaviors performed could best be described as behaviors more typically performed by children. In no way could the modeled behaviors (i.e., selecting toy hats, playing with toys) be described as "adult behavior." If, indeed, children are exposed to adults behaving like this, as in the preschool setting, they are equally exposed to children behaving the same way. The absence of a main effect for age of model indicates that children did not simply imitate adults more than children; rather, they imitated same-sex adults more.

In addition to examining children's actual behavior, an important component of the present study was the attention paid to models. That no sex differences were found indicates children did pay equal attention to male and female models or, at the very least, made no differentiation between them. Their correct recall of both males and females did not differ whether the models were adults or children. Thus, if children paid equal attention to both sexes, as evidenced by their recall scores, yet chose to enact mainly behaviors performed by members of their own sex, it suggests that children were not differentially processing gender-related information.

Because the only way children could be classified with adults in this instance is in terms of gender, the very fact that they paid greater attention to, and later imitated, adults suggests that same-sex adults are perceived as highly appropriate conveyors of in-group norms with respect to gender categories. However, because no condition in the present study required children to choose between adults and children directly, interpretations in this vein must be limited. This is addressed in Experiments 2 and 3.

Experiment 2

Although the results of Experiment 1 found no evidence for the superiority of peer, in preference to adult, influence, a criticism of that study was the inclusion of age of model only as a between-participants variable: Participants were not given the choice to model adults or children. The primary aim of Experiment 2, therefore, was to directly compare children's modeling of, and attention paid to, adults and other children. This was achieved by removing gender from the design: Child participants were shown videos of adult and child models, only of their own sex.

Another shortcoming of Experiment 1 was that the overall amount of modeling was relatively low, compared to the number of actions the participants were shown, and hence given the opportunity to imitate. Presumably, this was due to the difficulty of the task: Child participants viewed the entire behavior sequences and were then asked to perform the actions they saw without any prompts from the experimenter. Although this appeared to be the procedure used by Bussey and Bandura (1984), subsequent communication with the first author revealed that they had, in fact, shown the video in segments before the associated opportunity for imitative behavior. Hence, both the design of the study and the modeling task were simplified for Experiment 2.

Method

Participants. Participants were 17 boys and 17 girls attending a university-based child-care center in Canberra, Australia. Their ages ranged from 41 to 64 months (M = 54.21 months, SD = 7.07). As with Experiment 1, participants' parents were predominantly university staff and/or students from a range of ethnic backgrounds. Participants' parents provided written consent, and children provided verbal consent to child-care staff and to the experimenter.

Design. The study involved a 2 (sex of participant: boy, girl) × 2 (age of models: adult, child) mixed factorial design. Age of models was a within-participants factor, as children saw videos of adult and child models who were the same sex as themselves. The models in the videos were those from the previous experiment.

Procedure. As with Experiment 1, participants were tested individually. Two new behavior scripts were introduced in this study and two (counterbalanced) versions of each video were used. The behavior sequences in this study were comparable to those in Experiment 1: Models selected either a green or a white Mickey Mouse cap and, accordingly, enacted one of the two behavior sequences. As with Experiment 1, each behavioral sequence comprised 14 behavioral components.

After watching the video, children were again given the opportunity to model the behaviors they had seen. An edited version of the video was used, showing three or four actions of one adult and one child model at a time. The sequence of 14 behaviors was thus divided into five segments, with a break in the video to allow the participant to perform the preceding actions before showing the next behaviors. Behaviors were recorded on a checklist by the experimenter.

Following this, the same memory task used in Experiment 1 was administered. At the end of each session, children were thanked for their participation, allowed to choose a sticker to take home, and returned to the group play area.

Results

Modeled behavior. Although the video sequence contained 14 separate behaviors performed by each of the models, this number was reduced to 12 for coding purposes. One of the behaviors (a verbal statement) was omitted because the experimenter always clarified the left or right side with the participant when he or she was placing the sticker onto the pinboard. In addition, no participant said either of the final utterances, so this behavior was also omitted.

Hence, the remaining 12 behaviors were coded into one of the four categories (adult behavior, child behavior, new behavior, and no behavior performed).

The total number of behaviors in each of the four categories was summed. Due to the change in procedure from Experiment 1, the number of no behavior responses was much lower (20% of all modeling responses). Because of this, and consistent with Experiment 1, this category was omitted from analyses. Thus, a 2 (sex of participant) × 3 (modeled behavior: child, adult, new) mixed-model ANOVA, with modeled behavior as a within-participants factor, was performed.

This revealed a significant main effect for modeled behavior, $F(2, 64) = 33.09$, $p < .001$, partial $\eta^2 = .51$, indicating that children (both boys and girls) modeled other children ($M = 5.50$, $SD = 2.21$) significantly more than they modeled adults ($M = 3.38$, $SD = 2.35$), $t(33) = 3.02$, $p = .01$. Similarly, children's modeling of adults was significantly greater than their performance of new behaviors ($M = 0.71$, $SD = 1.45$), $t(33) = 4.86$, $p < .001$. Presumably due to the change in procedure for the modeling task, children's performance of a new behavior was extremely low in this study. Importantly, no sex differences were found; the patterns of modeling by boys and girls did not differ significantly from each other.

Consistent with Experiment 1, analyses were performed involving age of participants. Using a median split on children's age (median age = 54 months), a 2 (sex of participant) × 2 (age of participant) × 3 (modeled behaviors) mixed-model ANOVA was conducted. Again, there was a strong effect for modeled behaviors, $F(2, 60) = 32.11$, $p < .001$, partial $\eta^2 = .52$, along with a main effect for age of participant, $F(1, 30) = 4.86$, $p = .03$, partial $\eta^2 = .14$: Older children ($M = 3.39$, $SD = 0.48$) performed more behaviors than younger children ($M = 3.01$, $SD = 0.48$). No other main or interaction effects were found.

Memory for models. Questions concerning child behaviors and adult behaviors were each classified into three categories: correct, incorrect by substituting the opposite behavior, and other. The data were then analyzed in a 2 (sex of participant) × 4 (memory for models: child models correct, adult models correct, child models incorrect by adult behavior, and adult models incorrect by child behavior) mixed-model ANOVA, with memory for models as a within-participants factor.

The significant main effect for memory for models, $F(3, 96) = 12.65$, $p < .001$, partial $\eta^2 = .28$, merely revealed that participants' correct recall of behaviors (child models: $M = 5.18$, $SD = 2.05$; adult models: $M = 5.09$, $SD = 2.35$) was greater than their incorrect recall (child models: $M = 3.03$, $SD = 1.28$; adult models: $M = 2.89$, $SD = 1.34$). Age of participants was also investigated in a 2 (sex of participant) × 2 (age of participant) × 4 (memory) mixed-model ANOVA. No effect for age of participant was found. Only the main effect for memory was replicated, $F(3, 90) = 12.32$, $p < .001$, partial $\eta^2 = .29$.

Discussion

In this study, both boys and girls demonstrated a clear preference for imitating same-sex child models over same-sex adult models, and this occurred more than their tendency to perform an unseen behavior. No sex differences were found. Perhaps not surprisingly, older children performed more behaviors than did younger children. Older children are both more able to and more practiced at performing actions exhibited by other people.

Despite the strong preference for age-related imitation shown in the modeling task, memory for adult and child models did not differ. Moreover, children's correct recall was much greater than their tendency to misattribute behaviors to either the adult or the child models, which, in turn, was greater than their tendency to provide new responses. As was the case with gender categories in Experiment 1, children paid equal attention to models in both categories, yet chose to behave in accordance with the category most relevant to themselves. In this study, that category was age, and children modeled the behaviors of other children.

In contrast to Experiment 1, the results of the present study lend support to Harris's (1995) assertions. In a more explicit test of her claims, and removing gender as a possible confound, child participants imitated children more than they imitated adults. However, the removal of gender from the design of the present study actually changed the context from that of the first study. In the first study, the only available categorization was sex. Children overwhelmingly chose to imitate people (both adults and children) of their own sex. Although their imitation of same-sex adults was greater than that of children, their imitation based on gender was greater than that based on age. In the present study, where children only saw models who were the same sex as themselves (thereby eliminating any reason for categorization along gender lines), they imitated the child models (same-age imitation) to a greater extent than they imitated adult models. Such contextual changes, as occurred in these two studies, made it appropriate for children to act (i.e., to imitate) in accordance with their membership of the particular salient category:

Boys imitated males and girls imitated females in the first study, whereas children of both sexes imitated children, in preference to adults, in the second study.

These results provide additional evidence of young children's capacity to self-categorize according to either age or gender. This is similar to Sani and Bennett's (2001) study in which 6- and 7-year-olds varied their choice of stereotypical descriptive traits for groups depending on the comparative context in which their choice was made. Such contextual changes are reflected in the self-categorical choices made by the children—as demonstrated by children's descriptive terms (Sani & Bennett, 2001) or by behavioral measures such as imitation (in the current studies). Thus, if children are able to identify with, and therefore imitate, either same-sex people (regardless of age) or same-age people (regardless of sex) depending on which category (age or sex) was salient at the time, this could be demonstrated by experimentally manipulating (age and sex) category salience. This was explored in Experiment 3.

Experiment 3

There is abundant evidence demonstrating the effects of contextual changes on intergroup processes with adults (e.g., Haslam, Oakes, Turner, & McGarty, 1995; Platow, Mills, & Morrison, 2000). Considerably less research has been conducted with children, with most focusing on primary school–aged children. Notably, Bigler and colleagues (Bigler, 1995; Bigler, Brown, & Markell, 2001; Bigler et al., 1997; Patterson & Bigler, 2006) have shown that children as young as preschool age show in-group biases on the basis of novel, randomly assigned groups when these groups are made salient by authority figures via the labeling and use of the groups (though not when those groups are ignored) and increase their use of gender stereotyping as a result of teachers' labeling and use of gender in the classroom (Bigler, 1995). Similarly, Powlishta (1995a, 1995b, 2000) has demonstrated the importance of the intergroup context in children's judgments concerning gender. With the exception of Patterson and Bigler (2006), the youngest children in these studies were kindergarteners. Moreover, variations in context were only due to variations in the emphasis placed on gender in the different settings (similar to Experiments 1 and 2).

Our third experiment was therefore designed to vary the context (by manipulating both age and gender salience) in which children would be presented with the exact same information (similar to Ryan & David's, 2003, work with adults). Specifically, if age were made salient, it was hypothesized that children would imitate other children (in preference to adults and regardless of sex), but when gender was made salient, girls would imitate females (in preference to males and regardless of age) and boys would imitate males (in preference to females and regardless of age). All the while, however, children would be expected to pay equal attention to, and therefore be able to recall, both sets of information equally. This is because the behavioral choice concerning which behavior to adopt is, it is suggested by the previous two experiments, indeed a choice.

Method

Participants. Participants were 32 girls and 32 boys from four child-care centers in Canberra and Melbourne, Australia. Their ages ranged from 40 to 66 months (M = 50.5 months, SD = 7.64), and the children came from a variety of ethnic backgrounds. All participants had written parental consent and provided verbal consent themselves.

Design. The study involved a 2 (sex of participant) × 2 (category salience: age, gender) between-participants factorial design. Participants watched one of four videos depicting same-sex adult and opposite-sex child models (two videos were counterbalanced versions). The models and behavior sequences were those used in the previous experiments.

Procedure. Children were again tested individually in a separate area of the child-care center. Before watching the video, they were asked about activities that children and grown-ups (age salience condition) or males and females (gender salience condition) like to do. Specifically, they were asked, "What are some of the things that children (or ladies and girls) like to do?" followed by "What are some of the things that grown-ups (or boys and men) like to do?" They were then given a sorting task. Children were shown a series of 20 laminated color pictures, presented one at a time, which they were asked to sort into two piles: either children and adults (age salience condition) or males and females (gender salience condition).

Children were then asked to watch a video (involving same-sex adults and opposite-sex children) and given the opportunity for imitative behavior. The simplified modeling task from Experiment 2 was used, with the experimenter asking the child about his/her actions if the actions were unclear.

In an attempt to increase recall responses, a modified version of the memory task was administered. Rather than allowing for free recall, as had been the case in the previous studies, children were given a forced-choice format. This consisted of both behavioral options that had appeared in the video. In editing the videos, the final utterance when the models returned to their chair was deleted. Hence, participants were asked a total of 16 memory questions, 8 pertaining to each of the models. Again, the order of presentation of these questions was counterbalanced across participants.

Results

All children correctly sorted all 20 pictures into the correct piles according to the (age or gender) directions they had been given.

Modeled behavior. The eight behaviors were coded into one of the four categories used in Experiments 1 and 2, and a 2 (sex of participant) × 2 (category salience: age, gender) × 3 (modeled behavior: salient category consistent, salient category inconsistent, new behavior) mixed-model ANOVA, with modeled behavior as a within-participants factor, was conducted.

Only a main effect for modeling emerged, $F(2, 120) = 57.74$, $p < .001$, partial $\eta^2 = .49$. Children imitated models of the same salient category as themselves—whether age or gender ($M = 3.67$, $SD = 1.64$)—significantly more than they imitated models of the salient-inconsistent category ($M = 2.73$, $SD = 1.45$), $t(61) = 2.57$, $p = .03$, and their imitation of both salient-consistent and salient-inconsistent modeling occurred significantly more than their performance of a new (or unseen) behavior ($M = 0.63$, $SD = 0.86$), $t(61) = 11.29$, $p < .001$, and $t(61) = 9.67$, $p < .001$ (salience consistent vs. new behavior and salience inconsistent vs. new behavior, respectively). Importantly, there was no interaction between sex of participant and modeled behavior, $F(2, 120) = 0.89$, nor between category salience (i.e., age vs. gender) and modeled behavior, $F(2, 120) = 0.20$. Boys and girls did not differ in their modeling responses to the age and gender salience manipulations nor in their amount of modeling in the gender-salient and age-salient conditions.

Memory for modeled behavior. Memory for models was recorded as either correct or incorrect according to salience-consistent or salience-inconsistent category membership. A 2 (sex of participant) × 2 (category salience: age, gender) × 2 (salience consistency: consistent, inconsistent) × 2 (recall: correct, incorrect) mixed-model ANOVA was performed, where salience consistency and recall were within-participants factors.

A main effect for salience consistency, $F(1, 60) = 107.30$, $p < .001$, partial $\eta^2 = .63$, was qualified by a Salience Consistency × Recall interaction, $F(1, 60) = 4.81$, $p = .03$, partial $\eta^2 = .07$. Although children's correct recall of salience-consistent ($M = 5.22$, $SD = 1.96$) and salience-inconsistent ($M = 5.75$, $SD = 1.51$) behaviors was greater than their incorrect recall of either (consistent: $M = 2.27$, $SD = 1.66$; inconsistent: $M = 1.81$, $SD = 1.48$), this difference was greater for incorrect recall of behaviors. Importantly, there was no effect for sex of participant or for category salience.

Discussion

The results of this study were clear. Depending on whether age or gender was made salient, children imitated models of their in-group on the salient dimension. That is, when age was made salient, girls chose to imitate the children (who in this case were boys) rather than the adults (who were women). When gender was salient, the girls imitated the women rather than the boys. Similarly, boys imitated men in the gender-salient condition, and they imitated girls in the age-salient condition. This last finding is particularly interesting, given boys' resistance to imitating females, demonstrated in Experiment 1 (see also Bussey & Bandura, 1984; Fagot, 1985b; Lamb & Roopnarine, 1979; Maccoby, 1980).

Consistent with Experiments 1 and 2, no sex differences were found in children's memory for behaviors. Boys and girls did not differ in the attention they paid to, or their memory for, the behaviors of models of either category. Similarly, there were no differences according to the salience of either gender or age. Indeed, children correctly remembered almost all behaviors in all conditions. Despite the salience manipulation occurring prior to children viewing the behaviors, they nevertheless paid attention to both sets of models and were able to recall the behaviors of each. It appears that children then chose to perform the behaviors in accordance with the self-categorization that had arisen from the salience manipulation. In the current study, the age and gender manipulations helped make the children perceptually ready for information consistent with, and motivated them to act in accordance with, those categories. These manipulations also provided a framework for the behaviors. Not only did children begin to think of themselves as either children (in the age salience conditions) or as boys or girls (in the gender salience conditions), the behaviors they saw in the video could be appropriately divided along these lines.

We contend that central to imitating behaviors consistent with this categorical division was the children's self-definitions as members of one of these categories. Although we had no explicit measures of categorization in the present study, the results are clear. When age was salient, it made sense for children to self-categorize as children; when gender was salient, it made sense to self-categorize as either boys or girls.

The salience manipulations in the present study provided a context for the participants' behavior. The sorting task provided a way of dividing their world —into males and females or into adults and children. This mirrored situations that young children face every day. Boys and girls are provided with, for example, different clothes to wear and different toys to play with and are even shown different ways to behave. Similarly, there are different expectations, and permissible behaviors, for children and adults. This can be illustrated with the responses of children in the current study when asked what children and grown-ups like to do. Some of the things that children liked to do were "play with toys" and "eat lollies." Among the things grown-ups "like" to do were "driving," "cook the dinner," "drink wine," and "clean things." Young children are well aware of these differences and that they may be expected to change their behavior according to the social context. A girl may be expected to be quieter and less active than her brother but noisier and more active than her parents.

Perhaps the most important implication of the results of the current study, however, concerns the importance of gender per se to 3- and 4-year-olds. Taken together, the results of these three experiments suggest that there is nothing inherently important about gender that is responsible for the strong display of gender-typed preferences, such as boys' and girls' preferences for toys, dress, playmates, and so forth (see LaFreniere et al., 1984; Maccoby & Jacklin, 1987; Moller & Serbin, 1996; Powlishta et al., 1993; Thorne, 1986). Rather, we suggest that this gender divide is, first, partially due to the important role that gender is given by others and, second, that young children are sensitive, and responsive, to this. Remarkably in the present study, children did not differ in their display of category-appropriate behavior in the different contexts. Boys imitated just as many behaviors of men, when gender was salient, as they did of girls, when age was salient. The current series of studies has shown that children could, and did, behave in accordance with the relevant categorizations of gender or age. Moreover, in this final study, children could, and did, behave in accordance with whichever of these categories was salient at the time, despite being presented with the exact same information (i.e., the behavior sequences in the video presentation) in the different conditions.

Importantly, there were no gender differences exhibited: Both the boys and the girls were similarly sensitive to the change in context, indicating that a different behavior was appropriate. The change in context provided a different meaning for those behaviors. Wearing a red hat was male behavior in one context but child behavior in another. Behavior that had previously been devoid of context was made contextually meaningful. Hence, the exact same behavior was interpreted differently depending on the social context in which it was presented.

Clearly then, young children are capable of adapting their (imitative) behaviors in ways that reflect variations in the contexts in which those behaviors occur. All the while, however, they appear to pay attention to an array of information provided to them. This is particularly the case with information pertaining to gender and age—characteristics that are frequently made salient in Australian society via explicit labeling and implicit use.

General Discussion

The present set of three experiments sought to examine preschoolers' understanding of categories, with a specific focus on gender in categorical terms. Given recent recognition of the importance of the categorization process as a guide to children's behavior, the experiments drew heavily on existing, and substantive, theories of social categorization. In utilizing SIT, we specifically drew on SCT, given its explicit integration of the personal and the social self into a single theoretical perspective. Although these theories were not conceived of as developmental theories, the psychological processes outlined in each have, in recent years, been increasingly applied to the developmental domain, particularly to work with children (e.g., Bennett, Lyons, Sani, & Barrett, 1998; Bigler et al., 1997; Powlishta, 1995b; Yee & Brown, 1992). Indeed, the approach put forward by each of these theories has been recognized as ideally suited for the integration of developmental and social psychology (Bennett & Sani, 2004) and plays a substantial role in Bigler and Liben's (2006, 2007) DIT.

Deriving from the basic assumption that gender development involves social categorical understanding, we examined the categorical nature of gender with a specific focus on children's imitation of observed others and the attention paid to information presented. In all three experiments, children's imitative preferences were consistent with the contextual variations presented, despite demonstrating equal recall of behaviors performed by members of both categorical in-groups and out-groups. It was suggested, therefore, that behavioral enactment is the result of a deliberate choice, made by the child, concerning appropriate behavior—behavioral choice that is not simply a function of a child's similarity to models. Rather, as demonstrated in Experiment 3, it appears that children's behavioral choices reflect their (contextually sensitive) self-categorizations.

That these findings were replicated across the studies, and applied equally to boys and girls, can also be seen as evidence against an innate sex differences perspective. Learning to be a girl or a boy does not follow some preordained path. Learning about gender involves an understanding of the categorical nature of gender—being one of many girls or boys—and an understanding that categories, themselves, are flexible, changeable, and context dependent. Learning about gender involves an integration of social categorical information into one's personal self-definition. These processes are complex, yet they are utilized from an early age.

The current studies also represent an important step toward specifying conditions under which children are more likely to use gender as a guide to their behavior and as evidence of suggestions made by Bigler and Liben (2006) that children's use of categorical information is impacted by societal rules and mores. We can, for example, intensify or diminish the likelihood of children choosing to enact behaviors consistent with gender expectations by manipulating the salience of gender or that of an alternative category such as age.

We recognize, however, that these findings occurred in an individual setting. Future research could well be directed at replicating these results in a group setting, although the potential practical and administrative problems of such research cannot be underestimated.

Although derived from a different theoretical framework, the results of the current studies, and the theoretical approach taken, should not be seen as completely divergent from contemporary theories of gender development. In light of recent debate on the topic (cf. Bandura & Bussey, 2004; Martin, Ruble, & Szkrybalo, 2002, 2004), the current work can be seen as providing additional, and complementary, support toward a contemporary integrated approach to gender development. Indeed, the incorporation of social psychological theory extends this work through the articulation of causal mechanisms involved in learning about gender. That these are the same processes involved in other intergroup, or categorical, experiences (e.g., stereotyping, prejudice; cf. Bigler & Liben, 2006, 2007) is consistent with the notion that gender does not constitute a "special" category and, as such, does not require its own exclusive theory. Conversely, future research focusing on domain-general psychological processes affecting children's understanding of their social world should continue to prove increasingly fruitful.

References

Bandura, A., & Bussey, K. (2004). On broadening the cognitive, motivational, and sociostructural scope of theorizing about gender development and functioning: Comment on Martin, Ruble and Szkrybalo (2002). *Psychological Bulletin*, *130*, 691–701.

Barrett, M. (2000). *The development of national identity in childhood and adolescence*. Inaugural lecture presented at the University of Surrey, Guildford, UK. Retrieved December 2004 from http://www.surrey.ac.uk/Psychology/staff/papers/mb-inaugural.html

Bem, S. L. (1981). Gender schema theory: A cognitive account of sex typing. *Psychological Review*, *88*, 354–364.

Bem, S. L. (1984). Androgyny and gender schema theory: A conceptual and empirical integration. *Nebraska Symposium on Motivation*, *32*, 179–226.

Bennett, M., Lyons, E., Sani, F., & Barrett, M. (1998). Children's subjective identification with the group and in-group favouritism. *Developmental Psychology*, *34*, 902–909.

Bennett, M., & Sani, F. (2004). Introduction: Children and social identity. In M. Bennett & F. Sani (Eds.), *The development of the social self*. Hove, UK: Psychology Press.

Bigler, R. S. (1995). The role of classification skill in moderating environmental influences on children's gender stereotyping: A study of the functional use of gender in the classroom. *Child Development*, *66*, 1072–1087.

Bigler, R. S., Brown, C. S., & Markell, M. (2001). When groups are not created equal: Effects of group status on the formation of intergroup attitudes in children. *Child Development*, *72*, 1151–1162.

Bigler, R. S., Jones, L. C., & Lobliner, D. B. (1997). Social categorization and the formation of intergroup attitudes in children. *Child Development*, *68*, 530–543.

Bigler, R. S., & Liben, L. S. (2006). A developmental intergroup theory of social stereotypes and prejudice. In R. V. Kail (Ed.), *Advances in child development and behavior* (Vol. 34, pp. 39–89). San Diego, CA: Elsevier.

Bigler, R. S., & Liben, L. S. (2007). Developmental intergroup theory: Explaining and reducing children's social stereotyping and prejudice. *Current Directions in Psychological Science*, *16*, 162–166.

Brown, P. M., & Turner, J. C. (2002). The role of theories in the formation of stereotype content. In C. McGarty, V. Y. Yzerbyt, & R. Spears (Eds.), *Stereotypes as explanations: The formation of meaningful beliefs about social groups* (pp. 67–89). Cambridge, UK: Cambridge University Press.

Bussey, K., & Bandura, A. (1984). Influence of gender constancy and social power on sex-linked modeling. *Journal of Personality and Social Psychology*, *47*, 1292–1302.

Bussey, K., & Perry, D. G. (1982). Same-sex imitation: The avoidance of cross-sex models? *Sex Roles*, *8*, 773–784.

Cahill, B., & Adams, E. (1997). An exploratory study of early childhood teachers' attitudes toward gender roles. *Sex Roles, 36*, 517–529.

Condry, J., & Condry, S. (1976). Sex differences: A study of the eye of the beholder. *Child Development, 47*, 812–819.

Durkin, K. (1985). *Television, sex roles and children: A developmental social psychological account*. Milton Keynes, UK: Open University Press.

Fabes, R. A., Martin, C. L., & Hanish, L. D. (2003). Young children's play qualities in same-, other-, and mixed-sex peer groups. *Child Development, 74*, 921–932.

Fagot, B. I. (1985a). Beyond the reinforcement principle: Another step toward understanding sex-role development. *Developmental Psychology, 21*, 1097–1104.

Fagot, B. I. (1985b). The child's expectations of differences in adult male and female interactions. *Sex Roles, 11*, 593–600.

Feiring, C., & Lewis, M. (1987). The child's social network: Sex differences from three to six years. *Sex Roles, 17*, 621–636.

Frable, D. E. S., & Bem, S. L. (1985). If you are gender schematic, all members of the opposite sex look alike. *Journal of Personality and Social Psychology, 49*, 459–468.

Freud, S. (1924). *Collected papers*. New York: International Psychoanalytic Press.

Freud, S. (1933). *New introductory lectures on psychoanalysis* (J. Strachey, Trans.). New York: Norton.

Harris, J. R. (1995). Where is the child's environment? A group socialization theory of development. *Psychological Review, 102*, 458–489.

Harris, J. R. (1998). *The nurture assumption: Why children turn out the way they do*. London: Bloomsbury.

Haslam, S. A., Oakes, P. J., Turner, J. C., & McGarty, C. (1995). Social categorization and group homogeneity: Changes in the perceived applicability of stereotype content as a function of comparative context and trait favourableness. *British Journal of Social Psychology, 34*, 139–160.

Hewstone, M., Hantzi, A., & Johnston, L. (1991). Social categorization and person memory: The pervasiveness of race as an organizing principle. *European Journal of Social Psychology, 21*, 517–528.

Heyman, G. D., & Gelman, S. A. (2000). Preschool children's use of novel predicates to make inductive inferences about people. *Cognitive Development, 15*, 263–280.

Huston, A. C. (1983). Sex-typing. In E. M. Hetherington (Ed.), *Handbook of child psychology: Socialization, personality and social development* (Vol. 4, pp. 388–467). New York: Wiley.

Jackson, L. A., & Hymes, R. W. (1985). Gender and social categorization: Familiarity and ingroup polarization in recall and evaluation. *Journal of Social Psychology, 125*, 81–88.

Kohlberg, L. (1966). A cognitive-developmental analysis of children's sex-role concepts and attitudes. In E. E. Maccoby (Ed.), *The development of sex differences* (pp. 82–173). Stanford, CA: Stanford University Press.

LaFreniere, P., Strayer, F. F., & Gauthier, R. (1984). The emergence of same-sex affiliative preferences among preschool peers: A developmental/ethological perspective. *Child Development, 55*, 1958–1965.

Lamb, M. E., & Roopnarine, J. L. (1979). Peer influence on sex role development in preschoolers. *Child Development, 50*, 1219–1222.

Maccoby, E. E. (1980). *Social development: Psychological growth and the parent-child relationship*. New York: Harcourt, Brace, Jovanovich.

Maccoby, E. E. (1990). The role of gender identity and gender constancy in sex-differentiated development. *New Directions for Child Development, 47*, 5–20.

Maccoby, E. E., & Jacklin, C. N. (1987). Gender segregation in childhood. In H. W. Reese (Ed.), *Advances in child development and behavior* (Vol. 20, pp. 239–287). Orlando, FL: Academic Press.

Martin, C. L., & Fabes, R. A. (2001). The stability and consequences on young children's same-sex peer interactions. *Developmental Psychology, 37*, 431–446.

Martin, C. L., & Halverson, C. F. (1981). A schematic processing model of sex typing and stereotyping in children. *Child Development, 52*, 1119–1134.

Martin, C. L., & Halverson, C. F. (1983). The effects of sex-typing schemas on young children's memory. *Child Development, 54*, 563–574.

Martin, C. L., Ruble, D. N., & Szkrybalo, J. (2002). Cognitive theories of early gender development. *Psychological Bulletin, 128*, 903–933.

Martin, C. L., Ruble, D. N., & Szkrybalo, J. (2004). Recognizing the centrality of gender identity and stereotype knowledge in gender development and moving toward theoretical integration: Reply to Bandura and Bussey (2004). *Psychological Bulletin, 130*, 702–710.

McGarty, C., & Penny, R. E. C. (1988). Categorization, accentuation and social judgement. *British Journal of Social Psychology, 27*, 147–157.

McGarty, C., & Turner, J. C. (1992). The effects of categorization on social judgement. *British Journal of Social Psychology, 31*, 253–268.

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review, 100*, 254–278.

Moller, L. C., & Serbin, L. A. (1996). Antecedents of toddler gender segregation: Cognitive consonance, gender-typed toy preferences and behavioral compatibility. *Sex Roles, 35*, 445–460.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92*, 289–316.

Nesdale, D., & Flesser, D. (2001). Social identity and the development of children's group attitudes. *Child Development, 72*, 506–517.

Oakes, P. J., Turner, J. C., & Haslam, S. A. (1991). Perceiving people as group members: The role of fit in the salience of social categories. *British Journal of Social Psychology, 30*, 125–144.

Patterson, M. M., & Bigler, R. S. (2006). Preschool children's attention to environmental messages about groups: Social categorization and the origins of intergroup bias. *Child Development, 77*, 847–860.

Pitcher, E. G., & Schultz, L. H. (1983). *Boys and girls at play: The development of sex roles*. South Hadley, MA: Bergin and Garvey.

Platow, M. J., Mills, D., & Morrison, D. (2000). The effects of social context, source fairness, and perceived self-source similarity on social influence: A self-categorization analysis. *European Journal of Social Psychology, 30*, 69–81.

Powlishta, K. K. (1995a). Gender bias in children's perception of personality traits. *Sex Roles, 32*, 17–28.

Powlishta, K. K. (1995b). Intergroup processes in childhood: Social categorization and sex role development. *Developmental Psychology, 31*, 781–788.

Powlishta, K. K. (2000). The effect of target age on the activation of gender stereotypes. *Sex Roles, 42*, 271–282.

Powlishta, K. K., Serbin, L. A., & Moller, L. C. (1993). The stability of individual differences in gender-typing: Implications for understanding sex segregation. *Sex Roles, 29*, 723–737.

Rheingold, H. L., & Cook, K. V. (1975). The contents of boys' and girls' rooms as an index of parents' behavior. *Child Development, 46*, 445–463.

Rubin, J. Z., Provenzano, F. J., & Luria, Z. (1974). The eye of the beholder: Parents' views on sex of newborns. *American Journal of Orthopsychiatry, 44*, 512–519.

Ryan, M. K., & David, B. (2003). Gender differences in ways of knowing: The context dependence of the attitudes toward thinking and learning survey. *Sex Roles, 49*, 693–699.

Sani, F., & Bennett, M. (2001). Contextual variability in young children's gender ingroup stereotype. *Social Development, 10*, 221–229.

Serbin, L. A., Moller, L. C., Gulko, J., Powlishta, K. K., & Colburne, K. A. (1994). The emergence of gender segregation in toddler playgroups. In C. Leaper (Ed.), *Childhood gender segregation: Causes and consequences* (pp. 7–17). San Francisco: Jossey-Bass.

Serbin, L. A., Tonick, I. J., & Sternglanz, S. H. (1977). Shaping cooperative cross-sex play. *Child Development, 48*, 924–929.

Siegal, M. (1987). Are sons and daughters treated more differently by fathers than mothers? *Developmental Review, 13*, 147–183.

Tajfel, H., & Turner, J. C. (1986). The social identity theory of intergroup behavior. In S. Worschel & W. G. Austin (Eds.), *Psychology of intergroup relations* (2nd ed., pp. 7–29). Chicago: Nelson-Hall.

Taylor, S. E., Fiske, S. T., Etcoff, N. L., & Ruderman, A. J. (1978). Categorical and contextual bases of person memory and stereotyping. *Journal of Personality and Social Psychology, 36*, 778–793.

Thorne, B. (1986). Girls and boys together, but mostly apart. In W. W. Hartup & Z. Rubin (Eds.), *Relationships and development* (pp. 167-184). Hillsdale, NJ: Erlbaum.

Tittle, C. K. (1986). Gender research and education. *American Psychologist, 41*, 1161–1168.

Turner, J. C. (1984). Social identification and psychological group formation. In H. Tajfel (Ed.), *The social dimension: European developments in social psychology* (pp. 518-538). Cambridge, UK: Cambridge University Press.

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory*. Oxford, UK: Blackwell.

Verkuyten, M. (2002). Ethnic attitudes among minority and majority children: The role of ethnic identification, peer group victimisation and parents. *Social Development, 11*, 558–570.

Walker, P., & Antaki, C. (1986). Sexual orientation as a basis for categorization in recall. *British Journal of Social Psychology, 25*, 337–339.

Yee, M., & Brown, R. (1992). Self evaluations and intergroup attitudes in children aged three to nine. *Child Development, 63*, 619–629.

Yellond, N. (Ed.). (1998). *Gender in early childhood*. London: Routledge.

Yzerbyt, V. Y., Rocher, S., & Schadron, G. (1997). Stereotypes as explanations: A subjective essentialistic view of group perception. In R. Spears, P. J. Oakes, N. Ellemers, & S. A. Haslam (Eds.), *The psychology of stereotyping and group life* (pp. 20–50). Oxford, UK: Blackwell.

Child Development, November/December 2008, Volume 79, Number 6, Pages 1942–1956

Children's Responses to Interparental Conflict: A Meta-Analysis of Their Associations With Child Adjustment

Kimberly A. Rhoades
Stony Brook University

A meta-analysis examined the relations between children's adjustment and children's cognitive, affective, behavioral, and physiological responses to interparental conflict. Studies included children between 5 and 19 years of age. Moderate effect sizes were found for the associations between cognitions and internalizing and externalizing behavior problems and self-esteem problems, negative affect and behavioral responses and internalizing behavior problems, and behavioral responses and self-esteem problems. Small to moderate effect sizes were found for the associations between cognitions and relational problems, negative affect and behavioral responses and externalizing behavior problems, and physiological reactions and internalizing and externalizing behavior problems. Effect sizes were, with 1 exception, larger for internalizing than for externalizing behavior problems. Age significantly moderated the majority of effect sizes.

The link between interparental conflict (IPC) and children's behavioral and emotional dysfunction is well established in both intact and divorced families (Amato & Keith, 1991; Buehler et al., 1997; Cummings & Davies, 1994, 2002; Davies & Cummings, 1994; Emery, 1982; Grych & Fincham, 2001; Ingoldsby, Shaw, Owens, & Winslow, 1999). However, not all children who witness IPC display behavior problems (Jouriles, Murphy, & O'Leary, 1989). More recent research efforts have focused on identifying the characteristics of children exposed to IPC, their coping responses to conflict, and contextual factors of IPC that may relate to adjustment problems (e.g., Cummings & Davies, 2002).

Ultimately, it is not the conflict itself but rather more proximal processes that must account for the relation between IPC and child behavior problems. Children's responses to IPC are one possible proximal variable. Such responses are indicators of how children process and make meaning of IPC in relation to their own needs, desires, and goals. Of course, children's responses to IPC are not the only variables relevant to the association between IPC and child dysfunction. This meta-analysis focuses on children's responses to conflict because: (a) children's own responses to IPC are most proximal to their own psychosocial and physical adjustment; (b) these responses provide an index of how children interpret and cope with IPC, which should ultimately mediate the relation between IPC and child adjustment; and (c) the literature on children's reactions to IPC is sufficiently large to warrant a systematic, quantitative review and provides an established theoretical background. This meta-analysis focuses on four broad categories of child responses to IPC: cognitions, emotional responses, behavioral responses, and physiological responses. These four categories represent the full spectrum of possible child responses to conflict; when children are exposed to IPC they can think, feel, act, and physiologically respond to the situation. The first three of these responses to IPC are included as mediating constructs in three major theoretical frameworks. We first explore empirical research on the relations between cognitive, affective, behavioral, and physiological responses to IPC and child adjustment, drawing on relevant theoretical frameworks when warranted. We then outline what is still unclear in the literature and how a meta-analysis can advance knowledge in the field.

Cognitions

Hostile internal representations of IPC are consistently and strongly related to children's behavior problems (Davies, Forman, Rasi, & Stevens, 2002). The emotional security hypothesis (Davies & Cummings, 1994) posits that children's reactions to IPC are

---

Preparation of this article was supported by National Institute of Mental Health Grant 1R21MH068427. The author would like to thank Susan G. O'Leary, Tamara Del Vecchio, Carey Dowling, and Vincent Grande for their thoughtful reviews of the article and Erica Woodin for her assistance in coding the included articles for reliability purposes.

Correspondence concerning this article should be addressed to Kimberly A. Rhoades, Department of Psychology, Stony Brook University, Stony Brook, NY 11794-2500. Electronic mail may be sent to krhoades@notes.cc.sunysb.edu.

© 2008, Copyright the Author(s)
Journal Compilation © 2008, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2008/7906-0027

a function of the perceived implications of the conflict on the well-being of the family and have the goal of preserving and promoting the child's own emotional security. Thus, cognitions associated with threatened family security are likely to elicit fear and helplessness. These responses may, over time, become generalized responses to a variety of life events.

Additionally, researchers have found that cognitions of self-blame and threat are associated with internalizing behavior problems (Cummings, Davies, & Simpson, 1994; Grych, Fincham, Jouriles, & McDonald, 2000; Grych, Harold, & Miles, 2003; Grych, Seid, & Fincham, 1992; Jouriles, Spiller, Stephens, McDonald, & Swank, 2000; Kerig, 1998b; Rogers & Holmbeck, 1997). Consistent with the cognitive-contextual framework (Grych & Fincham, 1990; Grych et al., 2000), if children view their parents' conflicts as threatening to themselves or the family system or feel that they are unable to cope with the conflict they are likely to feel anxious and helpless. Likewise, if children feel that they are to blame for their parents' conflicts, they are likely to feel guilt, shame, and sadness (Grych & Fincham, 1990; Grych et al., 2000).

Children's cognitions could also be associated with aspects of development other than internalizing and externalizing behavior problems. For example, if children feel that they are to blame for their parents' conflict or that they are unable to cope with the conflict, they may be likely to have low self-esteem or self-worth (Rogers & Holmbeck, 1997). If children feel threatened by IPC, especially if they feel that the conflict may result in family dissolution or conflict spillover, their social relations within the family may be negatively influenced. Worry associated with threat cognitions or destructive family representations could also influence children's physical health through stress responses. Although we know relatively little about how children's cognitive responses to IPC influence their physical health, we do know that exposure to IPC is related to health problems (Gottman & Katz, 1989). Similarly, if children worry about how IPC may impact their family while they are in school, their academic progress may be influenced by their distraction. It is thus important to go beyond measures of internalizing and externalizing behavior problems when examining the possible impact of children's responses to IPC on child adjustment.

Affect

Child negative affect in response to IPC has also been investigated as a possible mediator of the relation between IPC and child adjustment. In response to IPC, sadness has been shown to be positively correlated with internalizing and externalizing behavior problems (Cummings, Goeke-Morey, & Papp, 2003). Fear has been shown to be positively correlated with internalizing behavior problems (Cummings et al., 2003; Crockenberg & Langrock, 2001). Anger, when coupled with paternal aggression, has been shown to be positively correlated with externalizing behavior problems in boys (Crockenberg & Langrock, 2001).

These findings are consistent with the cognitive-contextual framework (Grych & Fincham, 1990), the emotional security hypothesis (Cummings & Davies, 1994), and the specific emotions model (Crockenberg & Forgays, 1996; Crockenberg & Langrock, 2001). In the cognitive-contextual framework, emotions are conceptualized as primary responses to IPC, which then influence secondary, cognitive, processing of the event. Through its relation to cognitions, affect is thus viewed as a mediator of the relation between IPC and child adjustment. The emotional security hypothesis conceptualizes affect as an additional index of emotional security, which is ultimately related to child adjustment. Last, the specific emotions model posits that children's evaluations of IPC lead to affective reactions based on the children's expectations of specific goal attainment. These affective reactions are then theorized to relate to child internalizing and externalizing behavior problems.

Relations between affective reactions to IPC and child adjustment, however, have been inconsistent, especially for externalizing behavior problems. Although some authors find strong positive correlations between negative affect and internalizing (Davies & Cummings, 1998; Davies, Forman, et al., 2002) and externalizing (Ablow, 1997; Davies, Harold, et al., 2002) behavior problems, other authors have not found significant relations among these variables (Crockenberg & Langrock, 2001; El-Sheikh & Harger, 2001). It should be noted, however, that even nonsignificant correlations were positive; increased negative affect is certainly not associated with positive adjustment outcomes.

Behavioral Responses

Behavioral reactions to IPC are conceptualized as a final index of emotional security (Cummings & Davies, 1994) and are typically divided into involvement behaviors and avoidance behaviors. Involvement includes behaviors such as trying to break up the conflict or otherwise talking to or physically interacting with parents while they are arguing. Avoidance includes behaviors where children actively attempt to avoid their parents' conflict when they are aware of its occurrence. Behavioral reactions to IPC are not consistently positively related to adjustment problems. Although involvement in IPC is positively correlated with children's depression, low self-worth, anxiety, and hostility (O'Brien, Bahadur, Gee, Balto, & Erber, 1997; O'Brien, Margolin, & John, 1995), avoidance of IPC is negatively correlated with internalizing behavior problems in some studies (e.g., O'Brien et al., 1995) and positively correlated with internalizing behavior problems in other studies (e.g., Davies, Forman, et al., 2002). If a child avoids IPC and proceeds to ruminate over the conflict and or its implications for the family, he or she may be likely to develop internalizing behavior problems. Conversely, if a child avoids IPC by engaging in adaptive coping responses, such as distracting him- or herself with another positive activity, he or she might avoid the negative effects of IPC. Avoidance of IPC and externalizing behavior problems have not been found to be significantly correlated (see Davies, Forman, et al., 2002; O'Brien et al., 1995).

Physiological Responses

Physiological responses to IPC have been less frequently examined as possible mediators or moderators in the relation between IPC and child adjustment. Consistent with the emotional security hypothesis (Cummings & Davies, 1994), physiological responses to IPC may be indicators of the meaning placed on the conflict by the child. Thus, children's physiological responses to IPC may be associated with poor adjustment outcomes. Research has indicated that physiological responses to IPC are associated with adjustment outcomes (El-Sheikh, 2005; El-Sheikh, Harger, & Whitson, 2001; El-Sheikh & Whitson, 2006). Less is known, however, about the relative impact of physiological responses to IPC when compared to other child responses to IPC.

Age and Gender Effects

Inconsistent results have been found regarding the effects of gender on the relations between self-blame and threat cognitions and behavior problems. Some researchers find no gender differences (Grych et al., 1992; Jouriles et al., 2000). Others find that threat cognitions are a stronger predictor of dysfunction in boys, and self-blame cognitions are a stronger predictor of dysfunction in girls, even though boys and girls report similar rates of self-blame and threat cognitions (Cummings et al., 1994; Kerig, 1998b). Additionally, overall maladaptive cognitions were, in one study, a stronger predictor of externalizing behavior problems in boys and of internalizing behavior problems in girls (Kerig, 1998a).

Gender differences in the relations between negative affect and child adjustment have also been inconclusive. For example, El-Sheikh (2005) found that relations between self-reported anger and internalizing and externalizing behavior problems and self-reported sadness and externalizing behavior problems are stronger for girls than for boys. Additionally, Davies, Forman, et al. (2002) found gender moderation for links between cognitive and behavioral responses to IPC and maternal (but not child, paternal, or teacher) reports of child behavior problems. Other studies, however, have either failed to find gender differences (Cummings et al., 2006; Davies, Harold, et al., 2002) or have not directly examined the role of gender (Dukewich, 2001).

Child age could also be an important moderator of the relation between children's responses to IPC and their adjustment. Unfortunately, the role of child age is even less clear than that of gender. Most empirical studies use restricted age ranges or do not report analyses including child age as a predictor or covariate. Young children's behavior may be more highly influenced by primary processing (i.e., negative affective responses), whereas older children's behavior may be more highly influenced by secondary processing (i.e., cognitions; Grych & Fincham, 1990). We know that as children age, cognitions become less egocentric, less catastrophic about negative events, and more sophisticated (Harter, 1983; Leitenberg, Yost, & Carroll-Wilson, 1986; Piaget, 1932). Additionally, attributional cognitions become more highly correlated with depression as age increases (Nolen-Hoeksema, Girgus, & Seligman, 1992; Turner & Cole, 1994). Thus, as children age, they are less likely to make self-blaming (i.e., egocentric) or threat (i.e., catastrophic) attributions about events (Grych, 1998; Jouriles et al., 2000), although cognitions overall become more highly correlated with dysfunction.

Aims

The literature on children's responses to conflict is well developed and in many ways consistent. However, because no single study has included cognitive, affective, behavioral, and physiological responses to IPC in relation to broad child adjustment, it is difficult to interpret the relative importance or influence of each child response in relation to various adjustment indices. Similarly, no single study has included, or should be expected to include, all possible measures of child adjustment. By aggregating studies that include a variety of adjustment outcomes in relation to children's responses to IPC, we are able to make general conclusions about how these responses relate to adjustment as a whole. We are also then able to compare the relations between child responses to IPC and various measures of adjustment. The literature has also been unable to make conclusive statements about the possible moderating role of age or gender in the relations between children's responses to IPC and child adjustment. If the major theories are designed to apply to all children, then we should expect that all relevant responses to IPC are related to child adjustment for each age group and for both boys and girls. If this is not true, we may need to adjust our theories to reflect that processes may differ between older and younger children or boys and girls.

A meta-analysis is important at this time to (a) quantify, in one report, the past 15 or so years of research on the relation between children's responses to IPC and child adjustment; (b) compare the relative impact of each of these responses to IPC on child adjustment; (c) reduce uncertainty about gender effects in the relations between children's responses to IPC and adjustment; and (d) examine age effects in the relations between children's responses to IPC and adjustment. It is important for researchers, clinicians, and policy makers to have consolidated, easily accessible information about responses to IPC that are likely to be more or less deleterious to child development. It is similarly important for them to know if boys or girls or certain age groups are more or less likely to be negatively influenced by various responses to IPC.

Consequently, the first goal of the meta-analysis is to evaluate the research support for the relations between child adjustment and children's cognitive, affective, behavioral, and physiological responses to IPC. In relation to this goal, the relative effect sizes for the four constructs are compared with the goal of more fully understanding which, if any, of the constructs are more or less related to child adjustment. The second goal of the meta-analysis is to examine possible differences in the relations between children's responses to IPC and different types of adjustment difficulties. For instance, are children's responses to IPC more strongly related to internalizing behavior problems or externalizing behavior problems? The final goal is to evaluate the possible moderating effects of child age and gender for all significant relations between children's responses to IPC and child adjustment.

Method

To identify relevant studies for inclusion in the meta-analysis, computerized searches of five electronic databases—PsychInfo, PubMed, Dissertation Abstracts International, ERIC, and Sociological Abstracts—were conducted. Keywords used in the computerized search included interparent* conflict*, parent* conflict*, interparent* disagreement*, parent* disagreement*, child rearing disagreement*, child related disagreement*, child rearing conflict, child related conflict, behav* adjustment, coping, appraisal*, cognition*, reaction*, buffering factors, respon*, emotional response, behavioral response, child appraisal, adjustment, externalizing behavior problems, internalizing behavior problems, resilience, moderators, mediators, social development, domestic violence, and intimate partner violence (* indicates that any permutations of the word stem were retrieved). These keywords were selected based on a desire to include studies assessing a variety of IPC types, child responses to IPC, and adjustment variables in the meta-analysis. Studies including conflict about marital relations, children or childrearing practices, daily activities, and domestic violence were all included. No study was excluded based on type of IPC, type of child response to IPC, or type of adjustment variable. The search was limited to English language articles.

This initial search yielded a total of 7,422 studies (with expected overlap among the databases). To be included in the meta-analysis, each study met five criteria: (a) the study was published in English, (b) the study included at least one measure of child adjustment, and (c) the study included at least one measure of children's cognitive, emotional, behavioral, or physiological responses to IPC. Closer examination of the titles and abstracts of studies resulted in 130 studies that could possibly meet inclusion criteria. These studies were retrieved and more closely inspected. The reference lists of the retrieved articles, as well as all relevant meta-analyses, literature reviews, and books were then examined to identify any additional studies for inclusion. Nine journals likely to publish articles in this area:—*Journal of Family Psychology*, *Child Development*, *Journal of Abnormal Child Psychology*, *Journal of Marriage and Family*, *Developmental Psychology*, *Parenting: Science and Practice*, *Journal of Divorce and Remarriage*, *Journal of Clinical Child and Adolescent Psychology*, and *Journal of Child Psychology and Psychiatry*—were hand searched. This process yielded four additional studies. The above searches were conducted in September 2007. Finally, experts in the field were contacted to identify any relevant in press or unpublished studies. Experts in the field were identified as those researchers who were first authors on more than one published article in the area of children's responses to IPC and child adjustment. These 11 researchers were contacted once via e-mail. The combined literature search resulted in a total of 142 possibly relevant studies.

Of the 142 studies initially retrieved for inclusion in the meta-analysis, 17 did not include measures of child adjustment, 19 did not include measures of children's responses to IPC, 18 studies did not include the necessary statistical values, 7 articles included duplicate samples, and 16 articles were purely theoretical or reviews. The authors of the 18 studies that did not report the necessary statistical values were contacted via e-mail. Two authors responded with the necessary information. This resulted in a final group of 71 studies that were included in the current meta-analysis. These 71 studies came from 67 articles, with 4 articles containing two studies or data sets. The majority of included studies (52) were published journal articles; the remaining 19 were dissertations. Of these 71 studies, 50 included measures of cognitions, 18 included measures of negative affect, 23 included measures of behavioral reactions in response to IPC, and 6 studies included measures of physiological responses to IPC. Only 4 studies separated negative affect into specific emotions (Crockenberg & Langrock, 2001; Cummings et al., 2003; El-Sheikh, 2005; Harold, Shelton, Goeke-Morey, & Cummings, 2004). Because of this, analyses involving negative affect used composite measures of negative affect, averaging effect sizes for fear, anger, and sadness when they were reported separately. Cognitions were analyzed as a unified construct and separately for self-blame and threat cognitions. Behavioral reactions were analyzed both as a unified construct and separately for active avoidance of and involvement in IPC. Because of the heterogeneity of physiological responses in the literature, physiological responses to IPC could only be analyzed as a unified construct. Two of the included studies measured skin conductance, 1 measured skin conductance and heart rate, 1 measured cortisol reactivity, and 2 measured vagal regulation. Attempts were made to include the higher order construct of emotional security in the analyses. However, only two articles included the necessary zero-order correlations between overall emotional security and child adjustment. All articles that included relations between cognitive, affective, and behavioral components of emotional security were included in the review. Adjustment variables in the review include: internalizing and externalizing behavior problems, self-esteem problems, relational problems, academic or cognitive problems, physical health problems, emotion dysregulation, and poor self-esteem or self-worth. In most cases, there were not enough studies reporting a particular adjustment variable (the exceptions being internalizing and externalizing behavior problems, poor self-esteem, and relational problems) to separately meta-analyze the results. In these cases, the effect sizes for these adjustment variables were only included in the overall adjustment analyses.

Nine studies that included measures of cognitions reported results separately by gender. Thus, these studies were analyzed with both the entire sample and separately by gender. Because only a small number of studies reported results separately by gender, analyses for negative affect, behavioral reactions, and physiological responses were only run with the whole samples.

A very small number of studies that met inclusion criteria included longitudinal data. For consistency across studies, when the article included both cross-sectional and longitudinal data, the cross-sectional data were aggregated with the other cross-sectional studies. Because of the paucity of longitudinal data meeting inclusion criteria, longitudinal results were not aggregated but are later described in relation to aggregate cross-sectional findings.

Study Coding

Each of the 71 articles included in the meta-analysis were coded. Coded study characteristics included the type of publication, publication year, first author name, type of data collection method, mean child age, percentage of male children, percent minority and dominant minority group, clinical status of the population, percentage of married or cohabitating parents, and mean family income. Twenty-five percent of the articles were coded by an independent coder to determine reliability. Intraclass correlation coefficients for continuously coded variables ranged from .93 to 1.00. Cohen's kappa for categorically coded variables ranged from .88 to 1.00.

Results

All results are for weighted fixed effects analyses. A fixed effects model assumes that the effect size obtained in a study estimates the population effect sizes, with random error that derives solely from participant sampling (Lipsey & Wilson, 2001). This model is additionally used when one assumes that any heterogeneity in effect sizes is due to systematic differences between studies that can be modeled. Aggregate effect size and moderator analyses were conducted using SPSS Macros for Meta-Analysis Version 2005 (Wilson, 2005). After weighting for sample size, each effect size is given in correlation coefficient metric, ranging from −1 to +1. Effect sizes are positive when a higher degree of negative cognitions, negative affect, behavioral responses to IPC, or physiological responses to IPC was associated with more adjustment problems. Effect sizes are categorized as large, medium, and small if values were about equal to the following values: large = .40, medium = .25, and small = .10 (Lipsey & Wilson, 2001). For each significant aggregate effect size, two fixed effects analogue to regression analyses with age and methodological quality as potential moderating variables were conducted. Methodological quality was continuously scored with studies receiving 1 point each for using: (a) daily diary reports, (b) audio or video vignettes of IPC, and (c) observation of child behavior problems and/or child responses to IPC. Methodological quality ranged from 0 to 3 points with a mean of 0.18. For each significant effect size, fixed effects analogues to analyses of variance (ANOVAs) with publication status and dichotomous child age (above or below the mean) as the moderating variables were conducted. Additionally, for each significant aggregate effect size, the fail-safe *N* (Rosenthal, 1979) was computed. The fail-safe *N* estimates the number of studies reporting null findings needed to reduce the aggregate effect sizes to negligible levels. All fail-safe *N* analyses were conducted using the formula proposed by Orwin (1983). For this analysis, the critical effect size value was set at .01, representing a negligible effect size. Effect sizes were aggregated separately if at least three studies included the measure of child response to IPC or adjustment. Composite measures of overall maladaptive cognitions and overall behavioral responses to IPC were created to include all cognitive and behavioral variables, including those that were not reported in at least three studies. Similarly, overall adjustment indices were created to include all adjustment variables, including those that were not reported in at least three studies.

Descriptives

For this sample of 71 studies, the mean child age was 10.61 years (*SD* = 2.84), with a range of 5.0–19.3 years. Minority percentage ranged from 0% to 100% with a mean of 27.72% (*SD* = 23.90) minority participants. Of the studies that included minority participants, the dominant minority group was African American in 59.2% of the studies and Hispanic in 11.3% of the studies. The remaining studies including minority participants listed another minority group as the largest minority group or did not provide this information. Between 0% and 100% of the adult participants were married or cohabitating, with a mean of 92.46% (*SD* = 16.60). The mean yearly family income was $39,763.65 (*SD* = $18,781.95), with a range of $6,684–$88,000. About 74.6% of the studies had about equal numbers of boy and girl children, 11.3% had between 5% and 45% male children, 8.5% of the studies had 55% to 95% male children, and 1.4% had more than 95% male children. The remaining studies did not report this information. Forty-four percent of the studies used questionnaires as the sole data collection method, 7% used interviews as the sole data collection method, and 49% of the studies used a combination of methods.

Cognitions and Adjustment Problems

Fifty studies included information on the relations among cognitions about IPC and children's adjustment problems (Table 1). These studies included 10,364 participants, with a mean of 207.28 participants per study. Results are reported both for overall maladaptive cognitions (including self-blame cognitions, threat cognitions, control cognitions, and coping efficacy) and separately by self-blame and threat cognitions. The overall weighted aggregate effect size including all studies for the relation between cognitions and adjustment problems was $r = .18$, $p < .001$. Each of the more specific effect sizes was significant ($p < .001$), and most weighted aggregate effect sizes were medium-large (see Table 1 for values). The fail-safe *N* for these analyses ranged between 52 and 1,485. Each fail-safe *N* exceeded suggested minimum values (Rosenthal, 1984).

Nine of the 50 studies reported cognition results separately by gender. The 9 studies included 1,507 participants, with a mean of 167.44 participants per study. All weighted aggregate effect sizes were significantly ($p < .001$) different from zero. Most weighted aggregate effect sizes were of medium strength (see Table 1 for values). The fail-safe *N* indicated that between 152 and 384 unpublished studies with null results would be needed to reduce the aggregate effect sizes to negligible levels. Each fail-safe *N* exceeded suggested minimum values (Rosenthal, 1984).

Negative Affect and Adjustment Problems

Eighteen studies were included that measured relations between negative affect and adjustment problems (Table 2). These studies contained 3,538 total participants, with a mean of 196.56 participants per study. The overall weighted aggregate effect size including all studies for the relation between negative affect and adjustment problems was $r = .14$, $p < .001$. Each of the more specific weighted aggregate effect

Table 1
Cognitions and Adjustment Problems

Gender	Constructs	*N* studies	Effect size *r* (*SD*)	95% CI	*Z* score	Age moderation
Both	Cog and Int	45	.34 (0.18)***	.32 – .36	35.15	No
	Cog and Ext	43	.21 (0.14)***	.19 – .23	19.92	Yes
	Cog and Adj	50	.18 (0.10)***	.16 – .20	18.28	Yes
	SB and Int	37	.36 (0.13)***	.34 – .38	34.22	No
	SB and Ext	34	.28 (0.10)***	.25 – .30	24.15	Yes
	SB and Adj	39	.32 (0.09)***	.29 – .33	29.94	No
	Threat and Int	34	.40 (0.16)***	.38 – .42	37.44	Yes
	Threat and Ext	32	.21 (0.12)***	.19 – .23	18.05	Yes
	Threat and Adj	38	.30 (0.12)***	.28 – .32	27.88	Yes
	Cog and Rel	4	.14 (0.04)*	.02 – .27	2.22	No
	Cog and Self	3	.38 (0.03)***	.25 – .50	5.40	No
Boys	Cog and Int	9	.22 (0.10)***	.16 – .29	6.31	No
	Cog and Ext	9	.23 (0.10)***	.16 – .28	6.51	Yes
	SB and Int	9	.32 (0.22)***	.26 – .39	9.36	No
	SB and Ext	9	.21 (0.12)***	.15 – .28	6.08	Yes
	Threat and Int	7	.38 (0.12)***	.31 – .44	10.45	No
	Threat and Ext	7	.23 (0.10)***	.16 – .30	6.22	Yes
Girls	Cog and Int	8	.22 (0.06)***	.15 – .28	6.40	No
	Cog and Ext	8	.22(0.11)***	.16 – .28	6.49	Yes
	SB and Int	8	.31 (0.08)***	.25 – .37	9.32	No
	SB and Ext	8	.20 (0.08)***	.14 – .27	5.91	No
	Threat and Int	6	.36 (0.14)***	.30 – .42	10.59	Yes
	Threat and Ext	6	.25 (0.21)***	.18 – .31	7.04	Yes

Note. Cog = overall cognitions; SB = self-blame cognitions; Threat = threat cognitions; Adj = overall adjustment problems; Rel = relationship problems; Self = self-esteem problems; Int = internalizing behavior problems; Ext = externalizing behavior problems; Effect Size *r* = weighted aggregate effect size (correlation); *SD* = standard deviation of the weighted aggregate effect size; 95% CI = 95% confidence interval for the weighted effect size; Age Moderation = Is age a significant moderator of the effect size?
$^{*}p < .05$. $^{***}p < .001$.

sizes was significantly different ($p < .001$) from zero. Most weighted aggregate effect sizes were small to medium-large (see Table 2 for values). The fail-safe *N* indicated that between 224 and 480 unpublished null studies would be needed to reduce the aggregate effect sizes to negligible levels. Each fail-safe *N* exceeded suggested minimum values (Rosenthal, 1984).

Behavioral Responses to IPC and Adjustment Problems

Twenty-two studies were included that investigated the relations between behavioral responses to IPC and adjustment problems (Table 2). These studies included 5,618 total participants, with a mean of 255.36 participants per study. The overall weighted aggregate effect size including all studies for the relation between behavioral responses and adjustment problems was $r = .19$, $p < .001$. Each of the more specific weighted aggregate effect sizes, except one, was significantly ($p < .001$) different from zero. Weighted aggregate effect sizes were in the small–medium range (see Table 2 for values). The fail-safe *N* for these significant findings indicated that between 63 and 506 unpublished null findings would be required to reduce the aggregate effect sizes to negligible levels. Each fail-safe *N* exceeded suggested minimum values (Rosenthal, 1984). A nonsignificant ($p = .07$) weighted aggregate effect size was found for the relation between avoidance of IPC and externalizing behavior problems.

Physiological Responses to IPC and Behavior Problems

Six studies were included that investigated the relations among physiological responses to IPC and adjustment problems. These studies included 768 participants, with a mean of 128 participants per study. The overall weighted aggregate effect size including all studies for the relation between physiological responses and adjustment problems was $r = .12$, $p < .001$. Each of the more specific weighted aggregate effect sizes was significantly ($p < .001$) different from zero, and small in strength (see Table 2 for values). The fail-safe *N* indicated that between 60 and 78 studies with null findings would be required to reduce the aggregate effect sizes to negligible

Table 2
Negative Affect, Behavioral Responses, and Physiological Responses and Adjustment Problems

Constructs	*N* studies	Effect size *r* (*SD*)	95% CI	Z score	Age moderation
NA and Int	16	.31 (0.17)***	.28–.34	18.12	Yes
NA and Ext	16	.15 (0.13)***	.12–.19	7.89	Yes
NA and Adj	18	.14 (0.07)***	.11–.17	8.32	Yes
Beh and Int	22	.24 (0.19)***	.21–.26	17.55	Yes
Beh and Ext	22	.14 (0.13)***	.11–.17	9.07	Yes
Beh and Adj	22	.19 (0.13)***	.16–.22	13.74	Yes
Avoid and Int	10	.26 (0.26)***	.23–.30	15.31	Yes
Avoid and Ext	10	.04 (0.11)	.00–.09	1.82	Yes
Inv and Int	14	.29 (0.17)***	.26–.32	17.60	Yes
Inv and Ext	16	.15 (0.12)***	.12–.18	8.49	Yes
Beh and Self	3	.22 (0.06)**	.09–.35	3.20	No
Physio and Int	6	.14 (0.06)***	.07–.21	3.88	No
Physio and Ext	6	.11 (0.08)**	.04–.18	3.11	No
Physio and Adj	6	.12 (0.07)***	.05–.19	3.43	No

Note. NA = child negative affect; Beh = overall behavior response to interparental conflict; Avoid = avoidance of interparental conflict; Inv = involvement in interparental conflict; Physio = physiological response to interparental conflict; Adj = overall adjustment problems; Int = internalizing behavior problems; Ext = externalizing behavior problems; Self = self-esteem problems; Effect Size *r* = weighted aggregate effect size (correlation); *SD* = standard deviation of the weighted aggregate effect size; CI = 95% confidence interval for the weighted effect size; Age Moderation = Is age a significant moderator of the effect size?
**$p < .01$. ***$p < .001$.

levels. Each fail-safe *N* exceeded suggested minimum values (Rosenthal, 1984).

Effect Size Comparisons

Using r to z transformations, effect sizes for the relations between each of the four constructs and adjustment problems were compared. Aggregate effect sizes for the relations between cognitions and negative affect and internalizing behavior problems were significantly larger ($p < .05$) than those for behavioral and physiological responses to IPC and internalizing behavior problems (with one exception of negative affect and involvement in IPC). Aggregate effect sizes for the relations between cognitions and externalizing behavior problems were significantly larger ($p < .01$) than those for negative affect, behavioral responses, and physiological responses to IPC and externalizing behavior problems. Aggregate effect sizes for the relation between cognitions and self-esteem problems were significantly ($p < .001$) larger than the relation between behavioral responses to IPC and self-esteem problems. For overall adjustment, relations were significantly ($p < .001$) larger for cognitions and behavioral responses to IPC than for physiological responses to IPC.

Differences for Internalizing Compared to Externalizing Behavior Problems

Using r to z transformations, effect sizes for the relations between children's responses to IPC and internalizing behavior problems were compared to the effect sizes for the relations between children's responses to IPC and externalizing behavior problems. Other adjustment indices were not examined due to the small number of studies representing each adjustment outcome. With the exception of physiological responses, all effect sizes for the combined samples of boys and girls were larger for relations with internalizing behavior problems than for relations with externalizing behavior problems (all $p < .001$).

Age and Gender Moderation

For all significant effect sizes, analogue to regression moderation analyses were conducted with child age entered as the moderator. Child age was a significant moderator of 22 of 34 relations between child responses to IPC and adjustment problems (see Tables 1 and 2). In all instances, effect sizes were larger for older children than for younger children. To examine whether any of the effect sizes were nonsignificant for older children or younger children, analogue to ANOVA analyses were conducted with dichotomous child age as the moderator. Older children were defined as those at or above the mean of 10.61; younger children were defined as those below the mean of 10.61. Overall moderation results were equivalent to the previous regression results. Aggregate effect sizes were nonsignificant for the relations between overall cognitions and internalizing and externalizing behavior problems for younger girls, self-blame and externalizing behavior problems for younger boys, threat and internalizing behavior problems for younger girls, and threat and externalizing behavior problems for younger girls and younger boys ($p > .05$).

For those studies that reported effect sizes separately for boys and girls, effect sizes for the relations between self-blame and threat cognitions and adjustment for boys were compared to those for girls using r to z transformations. No significant differences were found (all $p > .05$).

Publication Status and Quality Moderation

For all significant effect sizes, analogues to ANOVA were conducted with publication status and methodological quality entered as moderators. Publication status moderated the magnitudes of the relations between 14 of 34 significant aggregate effect sizes (all *ps* $< .05$). In all cases, effect sizes were larger for published studies than for unpublished studies. Study quality moderated the magnitudes of the effect sizes for 11 of 34 significant aggregate effect sizes (all *ps* $<$ .05). Higher quality studies, for these relations, reported smaller effect sizes than lower quality studies.

Additional Analyses

Because the relations among children's reactions to IPC and their adjustment problems might be confounded with the amount of exposure children have to conflict, partial correlations controlling for the amount of IPC were computed for those studies that reported the necessary statistical values. The subsequent analyses are based on between 5 and 22 studies depending on the relation being analyzed. Partial aggregate effect sizes were significant for all relations with two exceptions: avoidance of IPC and internalizing and externalizing behavior problems (see Table 3). Bivariate aggregate effect sizes were significantly larger ($p < .05$) than partial aggregate effect sizes for all relations with the following exceptions: negative affect and externalizing behavior problems, avoidance and externalizing behavior problems, and involvement and externalizing behavior problems.

Although there were not enough longitudinal studies that met inclusion criteria to aggregate results across studies, bivariate correlations were extracted from these studies for comparison with cross-sectional aggregate results. Due to the small number of studies represented by these results, caution is urged in their interpretation. Two longitudinal studies provided bivariate longitudinal correlations among cognitions and adjustment problems, four studies (two within the same article) provided longitudinal correlations among negative affect and adjustment problems, five studies (two within the same article) provided longitudinal correlations among behavioral responses to IPC and adjustment problems, and three studies provided longitudinal correlations among physiological reactivity and adjustment problems. Longitudinal correlations were significantly lower than cross-sectional correlations for relations among cognitions and negative affect and internalizing behavior problems ($p < .01$).

Table 3
Partial Aggregate Effect Sizes Controlling for Amount of Interparental Conflict

Constructs	*N* studies	Effect size *r*	95% CI	Bivariate comparison
SB and Int	20	.30***	.28 – .33	Yes
SB and Ext	21	.17***	.15 – .20	Yes
Threat and Int	21	.36***	.34 – .38	Yes
Threat and Ext	22	.08***	.06 – .11	Yes
NA and Int	9	.24***	.18 – .29	Yes
NA and Ext	9	.16***	.10 – .22	No
Avoidance and Int	5	−.05	−.11 – .01	Yes
Avoidance and Ext	5	−.05	−.12 – .00	No
Involvement and Int	7	.06*	.01 – .11	Yes
Involvement and Ext	8	.10**	.05 – .15	No

Note. SB = self-blame cognitions; Threat = threat cognitions; Int = internalizing behavior problems; Ext = externalizing behavior problems; NA = negative affect; Effect Size *r* = weighted aggregate effect size (correlation); 95% CI = 95% confidence interval for the weighted effect size; Bivariate Comparison = Is the corresponding bivariate correlation significantly larger than the partial correlation?
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Discussion

The first goal of the meta-analysis was to evaluate and compare the research support for the relations among children's cognitions, negative affect, behavioral, and physiological reactions to IPC and their adjustment. Most aggregate effect sizes were significantly different from 0. Overall, the pattern of data implies that all four response constructs are related to child adjustment, although the relations appear stronger for cognitions and negative affect than for behavioral and physiological reactions to conflict.

Only one aggregate bivariate effect size was nonsignificant for the relation between avoidance and externalizing behavior problems. Avoidance is exclusively associated with internalizing behavior problems. If children are ruminating or worrying about the conflict during their avoidance, they may be likely to develop internalizing behavior problems. Alternatively, children who report more internalizing behavior problems may be more likely to avoid IPC in general, whereas children reporting externalizing behavior problems may be less likely to actively avoid IPC.

The second goal of the meta-analysis was to examine differences in the relations among children's responses to IPC and different types of adjustment problems. With the exception of involvement in IPC, relations between children's responses to IPC and internalizing behavior problems were larger than relations with externalizing behavior problems. If children use maladaptive coping strategies in response to IPC, they may be particularly vulnerable to using these same or similar maladaptive coping skills whenever they are confronted with difficult life circumstances. Inability to cope with various life stressors may make children particularly at risk for internalizing behavior problems. Because exposure to IPC and externalizing behavior problems are clearly associated, we need to know more about which, if any, child responses to IPC are more strongly or uniquely predictive of externalizing behavior problems. Although aggressive responses to IPC are related to general externalizing behavior problems (Cummings, Goeke-Morey, & Papp, 2004), the direction of this association is unclear.

The final goal of the meta-analysis was to examine the effects of age and gender on the relations between children's responses to IPC and their adjustment problems. Gender differences could only be explored for the relations among cognitions and internalizing and externalizing behavior problems. Interestingly, no significant differences were found between the aggregate effect sizes for girls and those for boys. This is consistent with previous studies that have failed to find gender differences in the relations between self-blame and threat cognitions and behavior problems (Grych et al., 1992; Jouriles et al., 2000). This does not imply that boys and girls are necessarily similarly susceptible to maladaptive cognitions; rather, these results imply that gender does not significantly influence how those cognitions are related to internalizing or externalizing behavior problems.

Age was found to be a significant moderator of the majority significant aggregate effect sizes. In each case, older children showed stronger associations between the two constructs than did younger children. Negative responses to IPC may be more likely to, over repeated experiences, become a characteristic response pattern to a variety of life events, resulting in more generalized internalizing or externalizing behavior problems. Effect sizes were nonsignificant for younger children for six of the relations between responses to IPC and adjustment. It may be that children younger than about age 10 lack the cognitive sophistication necessary for effectively generating and processing maladaptive cognitions about IPC. Younger children may also generate these cognitions while conflict is occurring but not continue to process them after the conflict resolves. Possibly, extended rumination about IPC increases one's risk of developing more generalized persistent behavior problems. Also, younger children have not had as much opportunity to witness IPC as older children. In line with the sensitization hypothesis (Davies, Myers, Cummings, & Heindel, 1999), conflict exposure may influence the relation among children's responses to IPC and their behavior problems. Although IPC exposure certainly influences child adjustment, aggregate partial correlations illustrate that children's responses to IPC (with the exception of avoidance) are significantly related to their adjustment even when exposure to IPC is statistically controlled. Particularly, in light of the nonsignificant effect sizes for the relations among some child responses to IPC and children's behavior problems for children under about 10 years, future studies should attempt to explain how and why relations between children's responses to IPC and their adjustment differ for elementary age versus adolescent children.

Result interpretation is limited by uncertainty about the directionality of the effects. Theories of the relations between children's responses to IPC and child adjustment have explained the process as being one in which children's reactions to IPC lead to child behavior problems (Davies & Cummings, 1994; Grych & Fincham, 1990). However, children's behavior problems could also influence their reactions to IPC. Children with internalizing behavior problems may be more likely to blame themselves for the conflict or feel threatened, anxious, fearful, or sad about the conflict, and be likely to avoid the conflict. Negative affect has been found to predict avoidant coping and be related to avoidance goals (Blackburn, Johnston, Blampied, Popp, & Kallen, 2006; Sideridis, 2005). Avoidance, then, could be the result of generalized internalizing behavior problems that manifest as fear or sadness in the presence of IPC than an antecedent of internalizing behavior problems.

Externalizing behavior problems could also influence children's reactions to IPC. For example, if children are generally aggressive, they are probably more likely to involve themselves in their parents' conflict than nonaggressive children. Externalizing behavior problems are indeed related to both aggressive responding to IPC in the home, and the endorsement of aggressive responses to IPC in laboratory analogue tasks (Cummings et al., 2004). Additionally, aggressive children are more likely to make hostile attributions (i.e., threat cognitions) than nonaggressive children (Dodge & Somberg, 1987). The relation between children's responses to IPC and their adjustment problems could thus be influenced by children's preexisting internalizing or externalizing behavior problems.

Longitudinal results, however, indicate that children's responses to IPC predict later behavior problems. For example, Cummings et al. (2006) found that IPC at Time 1 predicts 'children's responses to IPC at Time 2, which predicts internalizing and externalizing behavior problems at Time 3. Additionally, Schermerhorn, Cummings, DeCarlo, and Davies (2007) found that although behavioral dysregulation in response to IPC was positively associated with subsequent internalizing and externalizing behavior problems, internalizing and externalizing behavior problems were not significantly associated with later behavioral dysregulation. In other research (Grych et al., 2003; Schermerhorn, Cummings, & Davies, 2005), bidirectional effects are suggested. Although causality can only be definitively determined through experimental studies, carefully crafted longitudinal studies with multiple time points allow one to make strong inferences about causality and allows for the examination of the development of these processes. Harold et al. (2004) made advantageous use of this type of longitudinal design and were able to make causal assertions about the mediational role of cognitive, affective, and behavioral responses to IPC. In a positive move, longitudinal studies on this topic are becoming increasingly more common. It would be advantageous, however, to see longitudinal studies including all four types of child responses to IPC, and multiple measures of child adjustment, as well as additional analogue experimental studies.

Limitations of the meta-analysis should be noted. First, most of the included studies used questionnaires or structured interviews as the sole method of data collection. The accuracy with which responses on questionnaires or during interviews capture children's responses to IPC is unclear. Self-report methodology may overestimate true effect sizes, as studies using observation, daily diaries, and audio or video vignettes resulted in smaller aggregate effect sizes than studies using questionnaires. Although a small number of included studies used diary methods (see Cummings et al., 2003) or behavioral observations (see Davies & Cummings, 1998; El-Sheikh, 2005) to measure either children's responses to IPC or behavior problems, future studies could be improved by including these methods of data collection more frequently or by devising innovative methods for more accurately assessing children's responses to IPC. An additional limitation related to self-report methodology is that often children's responses to IPC and their behavior problems are reported by the same informant. This concern has been partially addressed by the nature of the literature. In many included reports, children's responses to IPC were reported by children, although their behavior problems were reported by their parents. Because this was not the case for all studies, and it could be argued that children's reports of internalizing behavior problems should always be used in lieu of parents' reports of these problems, the limitation cannot be completely addressed.

Sixteen studies were excluded because the authors did not include the necessary statistical values. The fail-safe *N*, however, indicates that even if these studies had null effect sizes, the results would remain significant. Higher order child responses to IPC could not be included in the meta-analysis due to statistical limitations. When higher order variables were included, relations between the latent variable and child adjustment were typically indexed multivariately. Because these studies controlled for different variables, we were unable to statistically aggregate findings related to higher order constructs.

A final limitation is that this meta-analysis focused solely on the relations among children's responses to IPC and child adjustment. This focus omitted studies that measured children's responses to IPC but did not include measures of child adjustment. It also omitted studies that focused on the relation between IPC and child adjustment but that did not include measures of children's responses to IPC. Ultimately, this meta-analysis was not designed to be a comprehensive review of all possible mediators or moderators of the relation between IPC and child adjustment, but rather a more focused review of the relations between one particular type of mediator, children's responses to IPC, and child adjustment.

Overall, the pattern of results indicates that children's cognitive, affective, behavioral, and physiological responses to IPC are moderately related to child adjustment. The results of this meta-analysis are informative in light of current theory about the mechanisms by which IPC influences children's behavior problems. It is clear that most child responses to IPC are significantly related to child adjustment. These results remain significant for all responses except avoidance even after controlling for conflict exposure. Results indicate that the cognitive-contextual framework, the emotional security hypothesis, and the specific emotions model all have something to contribute in explaining the relation between IPC and children's behavior problems. Because the cognitive-contextual framework and the emotional security hypothesis differ primarily by how they propose child responses to conflict influence adjustment and both, although to varying degrees, include cognitions, affect, and behavior in their models, it is difficult to evaluate which framework best describes the data. Cognitions and negative affect are more strongly related to adjustment problems than behavioral or physiological responses. However, all three theories include these two responses to IPC, making it difficult to argue for the supremacy of one theory over another. Because these theories include very similar constructs, it may be advantageous to further investigate the integration of the theories. Future studies should continue to concentrate on how each of these theories can inform one another and consider how different types of child responses may interact together in predicting child behavior problems.

References

References marked with an asterisk indicate studies included in the meta-analysis.

*Ablow, J. C. (1997). Marital conflict: Young children's perceptions and adjustment. *Dissertation Abstracts International, 59* (3B), 1356. (UMI No. 9828574)

Amato, P. R., & Keith, B. (1991). Consequences of parental divorce for the well-being of children: A meta-analysis. *Psychological Bulletin, 110*, 26–46.

*Anthony, C. M. (2000). Interparental conflict styles and youth problem behavior: The role of selected youth appraisal variables. *Dissertation Abstracts International: Section B. The Physical Sciences and Engineering, 61* (1-A), 381.

Blackburn, S., Johnston, L., Blampied, N., Popp, D., & Kallen, R. (2006). An application of escape theory to binge eating. *European Eating Disorder Review, 14*, 23–31.

*Blasey, C. M. (1997). Measuring young children's coping responses to interparental conflict. *Dissertation Abstracts International: Section B. The Physical Sciences and Engineering, 57* (9-B), 5943.

Buehler, C., Anthony, C., Krishnakumar, A., Stone, G., Gerard, J., & Pemberton, S. (1997). Interparental conflict and youth problem behaviors: A meta-analysis. *Journal of Child and Family Studies, 6*, 233–247.

*Buehler, C., Lange, G., & Franck, K. L. (2007). Adolescents' cognitive and emotional responses to marital hostility. *Child Development, 78*, 775–789.

*Burton, J. P. (1998). Marital conflict and child adjustment: Children's perceptions of marital conflict. *Dissertation Abstracts International: Section B. The Physical Sciences and Engineering, 59* (6-B), 3089.

*Cardoza-Fernandez, S. (2003). Investigating modifiable mediators of adjustment in children from violent homes: Implications for intervention. *Dissertation Abstracts International: Section B. The Physical Sciences and Engineering, 64* (5-B), 2380.

*Connell, A. M. (2004). Marital conflict and child adjustment problems: Links with children's social information processing and regulation of emotional arousal. *Dissertation Abstracts International: Section B. The Physical Sciences and Engineering, 64* (8-B), 4028.

*Crockenberg, S., & Forgays, D. K. (1996). The role of emotion in children's understanding and emotional reactions to marital conflict. *Merrill-Palmer Quarterly, 42*, 22–47.

*Crockenberg, S., & Langrock, A. (2001). The role of specific emotions in children's responses to interparental conflict: A test of the model. *Journal of Family Psychology, 15*, 163–182.

Cummings, E. M., & Davies, P. T. (1994). *Children and marital conflict*. New York: Guilford.

Cummings, E. M., & Davies, P. T. (2002). Effects of marital conflict on children: Recent advances and emerging themes in process-oriented research. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 43*, 31–63.

*Cummings, E. M., Davies, P. T., & Simpson, K. S. (1994). Marital conflict, gender, and children's appraisals and coping efficacy as mediators of child adjustment. *Journal of Family Psychology, 8*, 141–149.

*Cummings, E. M., Goeke-Morey, M. C., & Papp, L. M. (2003). Children's responses to everyday marital conflict tactics in the home. *Child Development, 74*, 1918–1929.

Cummings, E. M., Goeke-Morey, M. C., & Papp, L. M. (2004). Everyday marital conflict and child aggression. *Journal of Abnormal Child Psychology, 32*, 191–202.

*Cummings, E. M., Schermerhorn, A. C., Davies, P. T., Goeke-Morey, M. C., & Cummings, J. S. (2006). Interparental discord and child adjustment: Prospective investigations of emotional security as an explanatory mechanism. *Child Development, 77*, 132–152.

*Dadds, M. R., Atkinson, E., Turner, C., Blums, G. J., & Lendich, B. (1999). Family conflict and child adjustment: Evidence for a cognitive-contextual model of intergenerational transmission. *Journal of Family Psychology, 13*, 194–208.

*Davies, P. T. (1996). Children's emotional security as a mediator of the link between marital conflict and child adjustment. *Dissertation Abstracts International: Section B. The Physical Sciences and Engineering, 57* (6-B), 4053.

Davies, P. T., & Cummings, E. M. (1994). Marital conflict and child adjustment: An emotional security hypothesis. *Psychological Bulletin, 116*, 387–411.

*Davies, P. T., & Cummings, E. M. (1998). Exploring children's emotional security as a mediator of the link between marital relations and child adjustment. *Child Development, 69*, 124–139.

*Davies, P. T., Forman, E. M., Rasi, J. A., & Stevens, K. I. (2002). Assessing children's emotional security in the interparental relationship: The security in the interparental subsystem scales. *Child Development, 73*, 544–562.

*Davies, P. T., Harold, G. T., Goeke-Morey, M. C., Cummings, E. M., Shelton, K., & Rasi, J. A. (2002). Child emotional security and interparental conflict. *Monographs of the Society for Research in Child Development, 67*(3, Serial No. 270), 41–62.

Davies, P. T., Myers, R. L., Cummings, E. M., & Heindel, S. (1999). Adult conflict history and children's subsequent responses to conflict: An experimental test. *Journal of Family Psychology, 13*, 610–628.

*Davies, P. T., Sturge-Apple, M. L., Cicchetti, D., & Cummings, E. M. (2007). The role of child adrenocortical functioning in pathways between interparental conflict and child maladjustment. *Developmental Psychology, 43*, 918–930.

Dodge, K. A., & Somberg, D. R. (1987). Hostile attribution biases among aggressive boys are exacerbated under conditions of threat to self. *Child Development, 58*, 213–224.

*Dominguez, M. L. (1996). Children's perceptions of and reactions to marital conflict and violence. *Disserta-*

tion Abstracts International: Section B. The Physical Sciences and Engineering, 56 (9-B), 5194.

*Donaldson, D. L. (1996). Children's perceived control and coping strategies in response to interparental conflict. *Dissertation Abstracts International: Section B. The Physical Sciences and Engineering, 56* (10-B), 5762.

*Dukewich, T. L. (2001). Examining affect and cognition as mechanisms underlying the marital conflict-child adjustment relationship. *Dissertation Abstracts International: Section B. The Physical Sciences and Engineering, 62* (3-B), 1615.

*Dunn, J., O'Connor, T. G., & Cheng, H. (2005). Children's responses to conflict between their different parents: Mothers, stepfathers, nonresident fathers, and nonresident stepmothers. *Journal of Clinical Child and Adolescent Psychology, 34*, 223–234.

*El-Sheikh, M. (2005). The role of emotional responses and physiological reactivity in the marital conflict-child functioning link. *Journal of Child Psychology and Psychiatry, 46*, 1191–1199.

*El-Sheikh, M., Buckhalt, J. A., Cummings, E. M., & Keller, P. (2007). Sleep disruptions and emotional security are pathways of risk for children. *Journal of Child Psychology and Psychiatry, 48*, 88–96.

*El-Sheikh, M., & Harger, J. (2001). Appraisals of marital conflict and children's adjustment, health, and physiological reactivity. *Developmental Psychology, 37*, 875–885.

*El-Sheikh, M., Harger, J., & Whitson, S. (2001). Exposure to interparental conflict and children's adjustment and physical health: The moderating role of vagal tone. *Child Development, 72*, 1617–1636.

*El-Sheikh, M., Keller, P., & Erath, S. A. (2007). Marital conflict and risk for child maladjustment over time: Skin conductance level reactivity as a vulnerability factor. *Journal of Abnormal Child Psychology, 35*, 715–727.

*El-Sheikh, M., & Whitson, S. (2006). Longitudinal relations between marital conflict and child adjustment: Vagal regulation as a protective factor. *Journal of Family Psychology, 20*, 30–39.

Emery, R. E. (1982). Interparental conflict and the children of discord and divorce. *Psychological Bulletin, 92*, 310–330.

*Foo, G. B. (2002). Marital conflict and childhood relational aggression: A gendered-balanced approach to understanding adjustment problems in middle childhood. *Dissertation Abstracts International: Section B. The Physical Sciences and Engineering, 63* (6-B), 3042.

*Forman, E. M., & Davies, P. T. (2005). Assessing children's appraisals of security in the family system: The development of the Security in the Family System (SIFS) scales. *Journal of Child Psychology and Psychiatry, 48*, 900–916.

*Fosco, G. M., & Grych, J. H. (2007). Emotional expression in the family as a context for children's appraisals of interparental conflict. *Journal of Family Psychology, 21*, 248–258.

*Gerard, J. M., Buehler, C., Franck, K., & Anderson, O. (2005). In the eyes of the beholder: Cognitive appraisals as mediators of the association between Interparental conflict and youth maladjustment. *Journal of Family Psychology, 19*, 376–384.

*Gilliom, L. A. (1998). Children's appraisals of interparental conflict, perceived control and internalizing symptoms. *Dissertation Abstracts International: Section B. The Physical Sciences and Engineering, 58* (7-B), 3922.

Gottman, J. M., & Katz, L. F. (1989). Effects of marital discord on young children's peer interaction and health. *Developmental Psychology, 25*, 373–381.

Grych, J. H. (1998). Children's appraisals of interparental conflict: Situational and contextual influences. *Journal of Family Psychology, 12*, 1–17.

Grych, J. H., & Fincham, F. D. (1990). Marital conflict and children's adjustment: A cognitive-contextual framework. *Psychological Bulletin, 108*, 267–290.

Grych, J. H. & Fincham, F. D. (Eds.). (2001). *IPC and child development: Theory, research, and application*. New York: Cambridge University Press.

*Grych, J. H., Fincham, F. D., Jouriles, E. N., & McDonald, R. (2000). Interparental conflict and child adjustment: Testing the mediational role of appraisals in the cognitive-contextual framework. *Child Development, 71*, 1648–1661.

*Grych, J. H., Harold, G. T., & Miles, C. J. (2003). A prospective investigation of appraisals as mediators of the link between interparental conflict and child adjustment. *Child Development, 74*, 1176–1193.

*Grych, J. H., Jouriles, E. N., Swank, P. R., McDonald, R., & Norwood, W. D. (2000). Patterns of adjustment among children of battered women. *Journal of Consulting and Clinical Psychology, 68*, 84–94.

*Grych, J. H., Raynor, S. R., & Fosco, G. M. (2004). Family processes that shape the impact of interparental conflict on adolescents. *Development and Psychopathology, 16*, 649–665.

*Grych, J. H., Seid, M., & Fincham, F. D. (1992). Assessing marital conflict from the child's perspective: The children's perception of interparental conflict scale. *Child Development, 63*, 558–572.

*Harger, J. (2000). Children's physiological regulation and cognitive appraisals as moderators of the relationship between exposure to parental conflict and children's psychosocial functioning and physical health. *Dissertation Abstracts International: Section B. The Physical Sciences and Engineering, 60* (11-B), 5773.

*Harold, G. T., Fincham, F. D., Osborne, L. N., & Conger, R. D. (1997). Mom and dad are at it again: Adolescent perceptions of marital conflict and adolescent psychological distress. *Developmental Psychology, 33*, 333–350.

*Harold, G. T., Shelton, K. H., Goeke-Morey, M. C., & Cummings, E. M. (2004). Marital conflict, child emotional security about family relationships and child adjustment. *Social Development, 13*, 350–376.

Harter, S. (1983). Children's understanding of multiple emotions: A cognitive-developmental approach. In W. F. Overton (Ed.), *The relationship between social and cognitive development* (pp. 147–194). Hillsdale, NJ: Erlbaum.

*Ingoldsby, E. M., Shaw, D. S., Owens, E. B., & Winslow, E. B. (1999). A longitudinal study of interparental conflict, emotional and behavioral reactivity, and preschoolers' adjustment problems among low-income families. *Journal of Abnormal Child Psychology, 27*, 343–356.

Jouriles, E. N., Murphy, C. M., & O'Leary, K. D. (1989). Interspousal aggression, marital discord, and child problems. *Journal of Consulting and Clinical Psychology, 57*, 453–455.

*Jouriles, E. N., Spiller, L. C., Stephens, N., McDonald, R., & Swank, P. (2000). Variability in adjustment of children of battered women: The role of child appraisals of interparent conflict. *Cognitive Therapy and Research, 24*, 233–249.

*Kennedy, D. V. (2000). Effects of marital conflict on children's adjustment: The role of emotional regulation. *Dissertation Abstracts International: Section B. The Physical Sciences and Engineering, 60* (11-B), 5812.

Kerig, P. K. (1998a). Gender and appraisals as mediators of adjustment in children exposed to interparental violence. *Journal of Family Violence, 13*, 345–363.

*Kerig, P. K. (1998b). Moderators and mediators of the effects of interparental conflict on children's adjustment. *Journal of Abnormal Child Psychology, 26*, 199–212.

*Kilpatrick, K. L., & Williams, L. M. (1998). Potential mediators of post-traumatic stress disorder in child witnesses to domestic violence. *Child Abuse and Neglect, 22*, 319–330.

*Kitzmann, K. M., & Cohen, R. (2003). Parents' versus children's perceptions of interparental conflict as predictors of children's friendship quality. *Journal of Social and Personal Relationships, 20*, 689–700.

Leitenberg, H., Yost, L. W., & Carroll-Wilson, M. (1986). Negative cognitive errors in children: Questionnaire development, normative data, and comparisons between children with and without self-reported symptoms of depression, low self-esteem and evaluation anxiety. *Journal of Consulting and Clinical Psychology, 54*, 528–536.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

*Lutzke, J. (1996). Multiple mediation of the effects of interpersonal conflict on children's post-divorce symptomatology. *Dissertation Abstracts International: Section B. The Physical Sciences and Engineering, 56*(12-B), 7066.

*Mann, B. J., & Gilliom, L. A. (2004). Emotional security and cognitive appraisals mediate the relationship between parents' marital conflict and adjustment in older adolescents. *Journal of Genetic Psychology, 165*, 250–271.

*McDonald, R., & Grych, J. H. (2006). Young children's appraisals of interparental conflict: Measurement and links with adjustment problems. *Journal of Family Psychology, 20*, 88–99.

*Medina, A. M. (2003). The influence of family context on child variables: Examining linkages among interparental conflict, supportive parent-child relationships and children's appraisals, coping, and attention regulation. *Dissertation Abstracts International: Section B. The Physical Sciences and Engineering, 64* (6-B), 2929.

*Nicolotti, L., El-Sheikh, M., & Whitson, S. M. (2003). Children's coping with marital conflict and their adjustment and physical health: Vulnerability and protective functions. *Journal of Family Psychology, 17*, 315–326.

Nolen-Hoeksema, S., Girgus, J. S., & Seligman, M. E. P. (1992). Predictors and consequences of childhood depressive symptoms: A 5-year longitudinal study. *Journal of Abnormal Psychology, 101*, 405–422.

*O'Brien, M., Bahadur, M. A., Gee, C., Balto, K., & Erber, S. (1997). Child exposure to marital conflict and child coping responses as predictors of child adjustment. *Cognitive Therapy and Research, 21*, 39–59.

*O'Brien, M., Margolin, G., & John, R. S. (1995). Relation among marital conflict, child coping and child adjustment. *Journal of Clinical Child Psychology, 24*, 346–361.

*O'Connor, B. P., & Dvorak, T. (2002). Conditional associations between interparental conflict and adolescent problems: A search for personality-environment interactions. *Advances in Psychology Research, 14*, 213–237.

Orwin, R. G. (1983). The fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics, 8*, 157–159.

Piaget, J. (1932). *The moral judgment of the child*. New York: Harcourt Brace.

*Richmond, M. K., & Stocker, C. M. (2003). Siblings' differential experiences of marital conflict and differences in psychological adjustment. *Journal of Family Psychology, 17*, 339–350.

*Rodriguez, L. N., & Kitzmann, K. M. (2007). Coping as a mediator between interparental conflict and adolescents' romantic attachment. *Journal of Social and Personal Relationships, 24*, 423–439.

*Rogers, M. J., & Holmbeck, G. N. (1997). Effects of interparental aggression on children's adjustment: The moderating role of cognitive appraisal and coping. *Journal of Family Psychology, 11*, 125–130.

Rosenthal, R. (1979). The file drawer problems and tolerance for null results. *Psychological Bulletin, 86*, 638–641.

Rosenthal, R. (1984). *Meta-analytic procedures for social research*. Beverly Hills, CA: Sage.

Schermerhorn, A. C., Cummings, E. M., & Davies, P. T. (2005). Children's perceived agency in the context of marital conflict: Relations with marital conflict over time. *Merrill-Palmer Quarterly, 51*, 121–144.

Schermerhorn, A. C., Cummings, E. M., DeCarlo, C. A., & Davies, P. T. (2007). Children's influence in the marital relationship. *Journal of Family Psychology, 21*, 259–269.

*Schudlich, T. D. D., Shamir, H., & Cummings, E. M. (2004). Marital conflict, children's representations of family relationships, and children's dispositions toward peer conflict strategies. *Social Development, 13*, 171–192.

Sideridis, G. D. (2005). Goal orientation, academic achievement, and depression: Evidence in favor of a revised goal theory framework. *Journal of Educational Psychology, 97*, 366–375.

*Skopp, N. A., McDonald, R., Manke, B., & Jouriles, E. N. (2005). Siblings in domestically violent families: Experiences of interparent conflict and adjustment problems. *Journal of Family Psychology, 19*, 324–333.

*Stephens, N. S. (1995). Children of battered women: Mother/male partner conflict, child self-blame attributions and child behavior problems. *Dissertation Abstracts International: Section B. The Physical Sciences and Engineering, 55* (10-B), 4615.

*Stocker, C. M., Richmond, M. K., Low, S. M., Alexander, E. K., & Elias, N. M. (2003). Marital conflict and children's adjustment: Parental hostility and children's interpretations as mediators. *Social Development, 12*, 149–161.

*Tarnell, A. (2003). Children's perceptions of marital conflict and their self-esteem, peer relationships, and social competence. *Dissertation Abstracts International: Section B. The Physical Sciences and Engineering, 64* (4B), 1944.

*Tschann, J. M., Flores, E., Martin, B. V., Pasch, L. A., Baisch, E. M., & Wibbelsman, C. J. (2002). Interparental conflict and risk behaviors among Mexican American adolescents: A cognitive-emotional model. *Journal of Abnormal Child Psychology, 30*, 373–385.

Turner, J. E., & Cole, D. A. (1994). Developmental differences in cognitive diatheses for child depression. *Journal of Abnormal Child Psychology, 22*, 15–32.

*Ulu, I. P., & Fisiloglu, H. (2002). The relationship between Turkish children's perceptions of marital conflict and their internalizing and externalizing behavior problems. *International Journal of Psychology, 37*, 369–378.

Wilson, D. B. (2005). SPSS macros for meta-analysis [Computer software]. Retrieved March 2007 from http://mason.gmu.edu/~dwilsonb/ma.html

*Wynns, K. C. (2004). Marital conflict and adolescents' adjustment: Examining multiple protective factors. *Dissertation Abstracts International: Section B. The Physical Sciences and Engineering, 64* (9-B), 4643.

*Yuhav, R., & Sharlin, S. A. (2002). Blame and family conflict: Symptomatic children as scapegoats. *Child and Family Social Work, 7*, 91–98.

Manuscripts Accepted for Publication

Neonatal Stepping in Relation to Terrestrial Optic Flow, Marianne Barbu-Roth, David Anderson, Adeline Despres, Joelle Provasi, and Joseph Campos.

A Cognitive Approach to the Development of Early Language, Susan A. Rose, Judith F. Feldman, and Jeffery J. Jankowski.

Emergence of Global Shape Processing Continues Through Adolescence, K. Suzanne Scherf, Marlene Behrmann, Ruth Kimchi, and Beatriz Luna.

Face Adaptation and Attractiveness Aftereffects in 8-Year-Olds and Adults, Gizelle Anzures, Catherine Mondloch, and Christine Lackner.

Time Course of Visual Attention in Infant Categorization of Cats Versus Dogs: Evidence for a Head Bias as Revealed Through Eye Tracking, Paul C. Quinn, Matthew M. Doran, Jason E. Reiss, and James E. Hoffman.

Maternal Depression, Children's Attachment Security, and Representational Development: An Organizational Perspective, Sheree Toth, Dante Cicchetti, Fred A. Rogosch, and Melissa Sturge-Apple.

Group Nous and Social Exclusion: The Role of Theory of Social Mind, Multiple Classification Skill, and Experience of Peer Relations Within Groups, Dominic Abrams, Adam Rutland, Joseph Pelletier, and Jennifer M. Ferrell.

Age Differences in Future Orientation and Delay Discounting, Laurence Steinberg, Sandra Graham, Lia O'Brien, Jennifer Woolard, Elizabeth Cauffman, and Marie Banich.

Mother–Infant Vagal Regulation in the Face-to-Face Still-Face Paradigm in Moderated by Maternal Sensitivity, Ginger A. Moore, Ashley L. Hill-Soderlund, Cathi B. Propper, Susan D. Calkins, W. Roger Mills-Koonce, and Martha J. Cox.

How Adolescents Come to See Themselves as More Responsible Through Participation in Youth Programs, Dustin Wood, Reed Larson, and Jane Brown.

Children's Trust in Previously Inaccurate Informants Who Were Well or Poorly Informed: When Past Errors Can Be Excused, Erika Nurmsoo and Elizabeth Robinson.

Temperament, Harsh and Indulgent Parenting, and Chinese Children's Proactive and Reactive Aggression, Yiyuan Xu, Jo Ann M. Farver, and Zengxiu Zhang.

A Horse of a Different Color: Specifying With Precision Infants' Mappings of Novel Nouns and Adjectives, Amy Booth and Sandra R. Waxman.

The Development of Children's Ability to Use Evidence to Infer Reality Status, Ansley Tullos and Jacqueline D. Woolley.

Changes in Stress Perception and Coping During Adolescence: The Role of Situational and Personal Factors, Inge Seiffe-Krenke, Kaisa Aunola, and Jari-Erik Nurmi.

Adolescents' and Parents' Evaluations of Helping Versus Fulfilling Personal Desires in Family Situations, Judith Smetana, Marina Tasopoulos-Chan, Denise C. Gettman, Myriam Villalobos, Nicole Campione-Barr, and Aaron Metzger.

Individualizing Student Instruction Precisely: Effects of Child by Instruction Interactions on First Graders' Literacy Development, Carol Connor, Shayne B. Piasta, Stephanie Glasney, Barry Fishman, Christopher Schatschneider, Phyllis Underwood, Elizabeth Coyne Crowe, and Frederick J. Morrison.

Young Children's Reasoning About the Effects of Emotional and Physiological States on Academic Performance, Jennifer Amsterlaw, Kristin Hansen Lagattuta, and Andrew N. Meltzoff.

Social and Environmental Factors Associated With Preschoolers' Nonsedentary Physical Activity, William H. Brown, Karin A. Pfeiffer, Kerry L. McIver, Marsha Dowda, Cheryl L. Addy, and Russell R. Pate.

Pubertal Development: Correspondence Between Hormonal and Physical Development, Elizabeth Shirtcliff, Ronald Dahl, and Seth Pollak.

Resilience Among Children Exposed to Domestic Violence: the Role of Risk and Protective Factors, Cecilia Martinez-Torteya, G. Anne Bogat, Alexander von Eye, and Alytia A. Levendosky.

Harsh Parenting and Child Externalizing Behavior: Skin Conductance Level Reactivity as a Moderator, Stephen Erath, Mona El-Sheikh, and Mark Cummings.

Prenatal Cocaine Exposure and Infant Cortisol Reactivity, Rina Eiden, Yvette Veira, and Douglas A. Granger.

Maternal Insightfulness and Resolution of the Diagnosis Are Associated With Secure Attachment in Preschoolers With Autism Spectrum Disorder, David Oppenheim, Nina Koren-Karie, Smadar Dolev, and Nurit Yirmiya.

Children Who Read Words Accurately Despite Language Impairment: Who Are They and How Do They Do It? Dorothy Bishop, David McDonald, Sarah Bird, and Marianna Hayiou-Thomas.

Multiple Child Care Arrangements and Young Children's Behavioral Outcomes, Taryn Morrissey.

Boys Will Be Boys; Cows Will Be Cows: Children's Essentialist Reasoning About Gender Categories and Animal Species, Marianne G. Taylor, Marjorie Rhodes, and Susan Gelman.

Social Perception in Infancy: A Near Infrared Spectroscopy Study, Sarah Lloyd-Fox, Anna Blasi, Agnes Volein, Nick Everdell, Clare E Elwell, and Mark H. Johnson.

Probing the Neural Correlates of Anticipated Peer Evaluation in Adolescence, Amanda E. Guyer, Erin B. McClure-Tone, Nina D. Shiffrin, Daniel S. Pine, and Eric E. Nelson.

The Effects of Early Experience on Face Recognition: An Event-Related Potential Study of Institutionalized Children in Romania, Margaret C. Moulson, Alissa Westerlund, and Charles A. Nelson.

The Cognitive and Behavioral Characteristics of Children With Low Working Memory, Tracy Alloway, Susan Gathercole, Hannah Kirkwood, and Jullian Elliott.

Editorial Acknowledgments

Each calendar year we use the closing pages of the volume to offer thanks to the many people who help to make *Child Development* a success. This year's acknowledgments carry particular import because they are the final ones of my editorship, and thus it is a time to reflect not only on the preceding year, but also on my entire six-year editorial term.

As usual, I begin by conveying my gratitude to the colleagues named on the cover who have served as Associate Editors during the past year. I would like to single out for special recognition the six colleagues who began by the first issue of 2003 and then stayed with the journal until the very end of my term. Joe Allen, Marv Daehler, Kathy Hirsh-Pasek, Pat Miller, Sharon Nelson-Le Gall, and Paul Quinn have been editorial soul mates during the last six years, offering not only their superb scholarship, but also their steadfast support. I will always remain deeply grateful. Also essential to the review process are the Consulting Editors – whose names appear inside the cover—and the ad hoc reviewers—whose names are listed below. The editors, authors, and the discipline at large have benefited from their expert and dedicated work.

The Editorial Office in Ann Arbor also deserves our deep thanks. In particular, Detra Davis, the Managing Editor, has provided constant professional support, patience, and ever-present good cheer to editors, reviewers, and authors alike. Kaitlin (Kate) MacKenzie has continued to serve professionally as the Editorial Assistant. Also deserving special thanks are other key SRCD staff—Andrea Browning, Susan Lennon, and Mary Ann McCabe—and our Wiley-Blackwell production editor, Teresa Huang. Together they have made the operation of *Child Development* smooth and enjoyable. We thank all who have contributed during the last year and, indeed, during our entire editorial term.

—*Lynn S. Liben*

P. G. Aaron
J. Lawrence Aber
Jennifer Ablow
Frances Aboud
Adele Abrahamson
Linda Acredolo
Emma Adam
Lauren Adamson
Nikki Aikens
Nazan Aksan
David Amaral
Eric Amsel
Thomas Anders
Adam Anderson
David Anderson
Glenda Andrews
Mark Ansorge
Jason L. Anthony
Ian Apperly
Karen Appleyard
John Archer
Doreen Arcus
Jens Asendorpf
Steven Asher
Godwin Ashiabi
Cristina Atance
Marc Atkins
Martha Augoustinos
Catherine Bagwell
Lorraine Bahrick
Renee Baillargeon
Jodie Baird
Bruce Baker
Marian Bakermans-Kranenburg
Karen Bales
Brian Barber
Kim Bard
Yair Bar-Haim
Melissa Barnett
Steve Barnett
Rachel Barr
Helen Barrett
Marco Battaglia
George Bear
Theodore Beauchaine
Sarah Beck
Celia Beckett
Martha Ann Bell
Deborah Belle
Jay Belsky
Sheri Berenbaum
Sarah Berger
Lisa Berlin
Karen Bierman
Ann Bigelow
Dorothy Bishop
Maureen Black
Clancy Blair
Sarah Blakemore
Michel Boivin
Dorret Boomsma
Amy Booth
Julie Booth
Anne Borge
Heather Bortfeld
Michelle Bosquet
Anne Bowker
Julie Bowker
Robert Bradley
Charles Brainerd
Rebecca Brand
Stephen Brand
Julia Braungart-Rieker
Bethany Bray
Andrew Bremner
R. Brendgen
Patricia Brennan
Preston Britner
Darlene Brodeur
Danny Brom
Kathryn Brookmeyer
Marion Broome
Laurie Brotman
Brenda Bryant
Kristen Bub
Mandi Burnette
Melissa Burnham
Keith Burt
Emily Bushnell
Natasha Cabrera
Yvonne Caldera
Maureen Callanan
Nicole Campione-Barr

Linda Camras
John Capitano
Stephanie Carlson
Jeremy Carpendale
Thomas Carr
Victor Carrion
Alice Carter
Sue Carter
Leslie Carver
Juan Casas
Marianella Casasola
Jane Case-Smith
Avshalom Caspi
Margaret Caughy
Nicholas Cepeda
Bhismadev Chakrabarti
Tony Charman
Katarzyna Chawarska
Chuansheng Chen
Hsuan-Chih Chen
Xinyin Chen
Jane Childers
Michelle Chouinard
Antonius Cillessen
Andrei Cimpian
K. Clarke-Stewart
Christopher Coe
Nancy Cohen
Robert Cohen
Susan Cohen Levine
David Coker
Kevin Cokley
Jeffrey Coldren
David Cole
John Coley
Katherine Conger
Joan Cook
Shauna Cooper
Daniela Corbetta
Sara Cordes
Penny Corkum
Kai Cortina
Eugenia Costa-Giomi
Catherine Costigan
Mary Courage
Nelson Cowan
Philip Cowan
Martha Cox
Sarah Coyne
Michael Criss
Patricia Crittenden
Lisa Crockett
Eveline Crone
Tim Croudace
David Crystal
Anne Culp
Mark Cummings
William A. Cunningham
Fiery Cushman
Chris Daddis
Kim Dalton
Andrea Danese
Barbara Danis
Richard De Lisi
Marc de Rosnay
Scania de Schonen
Helene Deacon
Eric Dearing
Ganie DeHart
Judy DeLoache
Yvon Delville
Jack Demick
Daneen Deptula
Raye-Ann deRegnier
Kathryn Dewar
Karen Diamond
Rose DiBiase
David Dickinson
Cheryl Dissanayake
Theodore Dix
Wallace Dixon
Jennifer Dobbs
Kenneth Dodge
Wilberta Donovan
Bonny Donzella
Willa Doswell
Eric Dubow
Brad Duchaine
Greg Duncan
George DuPaul
Davido Dupree
Tina Du Rocher Schudlich
Jodi Dworkin
Matthew Dye
Diane Early
Warren Eaton
Catharine Echols
Christina Edmonds
Carolyn Edwards
Natalie Eggum
Linnea Ehri
Rebecca Eilers
Shiri Einav
Ann Eisenberg
Nancy Eisenberg
Thalia Eley
Bruce Ellis
Mona El-Sheikh
Karen Emmorey
Susan Engel
Michelle Englund
Jeffrey Epstein
Stephen Erath
Cynthia Erdley
David Estell
Richard Fabes
Jacqueline Fagard
Richard Faldowski
John Fantuzzo
Thomas Farmer
JoAnn Farver
Ruth Feldman
Deborah Feltz
Christopher Fennell
Anne Fernald
Russ Fernald
Linda Fetters
Tiffany Field
Barbara Fiese
William Fifer
Shalom Fisch
Cynthia Fisher
Phil Fisher
Paula Fite
Stanka Fitneva
Robyn Fivush
Connie Flanagan
Erika Forbes
Michael Foster
Elise Frank Masur
Janet Frick
Ori Friedman
Margaret Friend
Jan Frijters
Douglas Frye
Michael Furlong
Nancy Galambos
Alison Garton
Suzanne Gaskins
Sue Gathercole
Lisa Gatzke-Kopp
Jane Gaultney
Heidi Gazelle
Susan Gelman
Carol George
Peter Gerhardstein
LouAnn Gerken
Elizabeth Gershoff
Nancy Getchell
Simona Ghetti
Chandra Ghosh
John Gibbs
Christina Gibson-Davis
Jay Giedd
Jeffrey Gilger
Rick Gilmore
Tracy Gleason
Sharon Glennen
Jennifer Glick
Wendy Goldberg
Eugene Goldfield
David Goldman
Sara Goldstein
Nancy Gonzales
Beth Goodlin-Jones
Gail Goodman
Noah Goodman
Ian Goodyer
Alexandra Gottardo
Julia Graber
James Graham
Jonathan Green
Patricia Greenfield
Andrea Greenhoot
Marissa Greif
Kalanit Grill-Spector
Emily Grossman
Harold Grotevant
Betsy Groves
Joan Grusec
Steven Guberman
Antoine Guedeney

Michaela Gummerum
Guang Guo
Gwen Gustafson
Grant Gutheil
Amanda Guyer
Hyowon Gweon
Ann Hagell
David Haley
Elizabeth Hammock
Wen-Jui Han
Brian Hare
Gordon Harold
Lauren Harris
Paul Harris
Peter Hauser
Patricia Hawley
Harlene Hayne
Andrea Heberlein
Annette Henderson
Heather Henderson
Daphne Hernandez
Irva Hertz-Picciotto
David Hessl
Barbara Heyns
Jari Hietanen
Robert Hinde
Melissa Hines
Connie S. H. Ho
John Hobcraft
Ernest Hodges
Jill Hohenstein
Barbara Hohle
Rene Hoksbergen
George Holden
Bruce Homer
Marc Hooghe
Brian Hopkins
Rebecca Hoss
Nina Howe
Hui-Chin Hsu
Bin Huang
Julie Hubbard
Carol Huntsinger
Janet Hyde
Marilou Hyson
William Idsardi
Mutsumi Imai
Kayoko Inagaki
David Ingram
Atsushi Iriki
Moshe Israelashvili
Jana Iverson
Sandra Jacobson
Laudan Jahromi
Vikram Jaswal
Elizabeth Johnson
Kathy Johnson
Mark Johnson
Susan Johnson
Nancy Jordan
Sharon Judge
Vonda Jump
Laura Justice
Jerome Kagan
Kirsten Kainz
Nina Kaiser
Chuck Kalish
Tomoe Kanaya
Kathleen Kannass
Nancy Kanwisher
Katherine Karraker
Joan Kaufman
Robert Kavanaugh
Yoshito Kawabata
Kerry Kazura
Julian Keenan
Kate Keenan
Heidi Keller
Meret Keller
Monika Keller
Peggy Keller
Elizabeth Kelley
Diann Kelly
Nenagh Kemp
Boaz Keysar
Lisa Kiang
Julia Kim-Cohen
Thomas Kindermann
David Klahr
Wendy Kliewer
Ariel Knafo
George Knight
Joshua Knobe
Karestan Koenen
Melissa Koenig
Claire Kopp
Jana Kreppner
Patricia Kuhl
Valerie Kuhlmeier
Robert Kumsta
Gabriel Kuperminc
Tamar Kushnir
Kristin Lagattuta
Deborah Laible
Robert Laird
Michael Lamb
Sharon Lambert
Karin Landerl
Daniel Lapsley
Marta Laupa
Mathieu Le Corre
Amy Learmonth
Michele Lease
Amy Lederberg
Kang Lee
Eva Lefkowitz
Richard LeGrand
Elyse Lehman
Melinda Leidy
Patrick Leman
Kathryn Lemery
Liliana Lengua
Laurence Leonard
Leslie Leve
Tama Leventhal
Robert LeVine
Gary Levy
Marc Lewis
Marva Lewis
Michael Lewis
David Lewkowicz
Robert Lickliter
Angeline Lillard
David Liu
David Livesey
John Lochman
Tania Lombrozo
Matthew Longo
Michael Lopez
Catherine Lord
Maryanne Loughry
Beatriz Luna
Brenda Lundy
S. Lupien
Michael Lynch
Thomas Lyon
Karlen Lyons-Ruth
Lili Ma
Charles MacArthur
Kim MacLean
Katherine Magnuson
Ramaswami Mahalingam
Francoise Maheu
Tina Malti
Virginia Marchman
Stuart Marcovitch
Henry Markovits
Carol Martin
Craig Mason
Barbara Maughan
Daphne Maurer
Rachel Mayberry
Jessica Maye
Lara Mayeux
Ashley Maynard
Devin McAuley
Lorraine McCune
Jessica McDermott Sales
David McDowell
Nancy McElwain
Shirley McGuire
Clark McKown
Thomas McMahon
Bob McMurray
Nicole McNeil
James McPartland
Gerald McRoberts
Jacques Mehler
Elizabeth Meins
Roberto Mejia
Edward Melhuish
Sterett Mercer
Daniel Messinger
George Michel
Amori Mikami
Norweeta Milburn
Jonathan Mill
Alison Miller
Kevin Miller
Paul Miller
Scott Miller
Candice Mills

Nancy Minshew
Rashmita Mistry
Karen Mitchell
Stephen Mitroff
Kelly Mix
Peter Molenaar
Victoria Molfese
Catherine Mondloch
Catherine Monk
Christine Moon
Chris Moore
Holly Moore
Michael Morales
Amanda Morris
Barbara Morrongiello
Jeylan Mortimer
Dirk Moser
David Moshman
Frosso Motti-Stefanidi
Margaret Moulson
Ulrich Müeller
Eduard Mulder
Peter Mundy
Lynne Murray
Dianna Murray-Close
Daniel Nagin
Kristin Neff
Sonya Negriff
Jenae Neiderhiser
Charles Nelson
Eric Nelson
Keith Nelson
Lauren Nelson
Denise Newman
Marie-Pascale Noel
Illene Noppe
Larry Nucci
Jari-Erik Nurmi
Lisa Oakes
Jelena Obradovic
Margaret O'Brien
Thomas Olino
Bonamy Oliver
Tjeert Olthof
Stacy Overstreet
Elizabeth Owens
Judith Owens
Sally Ozonoff
Barbara Pan
Yingqiu Pan
Robin Panneton
Daniel Paquette
Oliver Pascalis
Helen Patrick
Tomas Paus
Vassilis Pavlopoulos
Katherine Pears
Liz Pellicano
Koraly Perez-Edgar
Susan Perlman
Josef Perner
Jennifer Pfeifer
Anne Pick
Andrew Pickles
Daniel Pine
Giuliana Pinto
Joe Piven
Robert Plomin
Linda Polka
Seth Pollak
Eva Pomerantz
Diane Poulin-Dubois
Thomas Power
Kimberly Powlishta
Sandeep Prasada
Patricia Prelock
Hazel Prelow
Joseph Price
Laura Proctor
Robert Provine
Raija-Leena Punamaki
Jodi Quas
Marcela Raffaelli
David Rakison
Beverley Raphael
Rebecca Ray
Bethany Reeb
Elaine Reese
Matthias Reitzle
Rena Repetti
Leslie Rescorla
Valerie Reyna
Chandra Reynolds
Soo Rhee
John Richards
Rebekah Richert
Carolien Rieffe
Deborah Rivas-Drake
Jenny Roberts
Scott Robinson
Phillip Rodkin
Robert Roeser
Sally Rogers
Mark Roosa
Linda Rose-Krasnor
Karl Rosengren
Mary Rothbart
Meredith Rowe
Stephanie Rowley
Penny Roy
Julie Rusby
James Russell
Adam Rutland
Gene Sackett
Avi Sadeh
Javid Sadr
Abraham Sagi-Schwartz
Arnold Sameroff
Dana Samson
Curt Sandman
Lynn Santelmann
Barbara Sarnecka
Megan Saylor
Suzy Scherf
Bambi Schieffelin
Louis Schmidt
Mary Schneider
Wolfgang Schneider
Sarah Schoppe-Sullivan
Laura Schulz
Lisa Scott
Stephen Scott
Matthew Scullin
Margretta Seashore
Eleanor Seaton
Amanda Seed
Amanda Seidl
Ronald Seifer
Monique Sénéchal
Atsushi Senju
Maria Sera
Joan Sereno
Simone Shamay-Tsoory
Timothy Shanahan
Daniel Shaw
Elizabeth Shirtcliff
Rainer Silbereisen
Wendy Silverman
Leher Singh
Wai Ting Siok
Cheryl Sisk
Virginia Slaughter
Cheryl Slomkowski
Anne Smith
Cynthia Smith
Lisa Smithers
Jesse Snedeker
Nancy Snidman
Catherine Snow
Margaret Snowling
Beate Sodian
Mary Solanto
Gregg Solomon
Jessica Sommerville
Edmund Sonuga-Barke
Victoria Southgate
Gottfried Spangler
Melanie Spence
John Spencer
Dan Sperber
Tracy Spinrad
Leyla Stambaugh
Kathy Stansbury
Håkan Stattin
Hans Steiner
Joan Stevenson-Hinde
Ian St. James-Roberts
Diane St. Laurent
Beth Stormshak
Deryn Strange
Tricia Striano
Eugene Subbotsky
Margaret Sullivan
Terri Sullivan
H. Swanson
Megan Sweeney
Daniel Swingley
Catherine Tamis-LeMonda
Li Hai Tan
Akaysha Tang
Twila Tardif
Louisa Tarullo

Eric Taylor
Lorraine C. Taylor
Marianne Taylor
Andrew Terranova
Douglas Teti
Donna Thal
Anita Thapar
David Thomas
Kathleen Thomas
G. Brian Thompson
John Tisak
Kathryn Tout
Dominique Treboux
Christopher Trentacosta
Penelope Trickett
Lara Triona
Georgene Troseth
John Trueswell
Jonathan Tubman
Elliot Turiel
Rhiannon Turner
Yuuko Uchikoshi
Rudolf Uher
Beverly Ulrich
Donald Unger
Ximena Uribe-Zarain
Kerstin Uvnäs-Moberg
Fons van de Vijver
Audrey Van der Meer
Patricia Van Horn
Roberto Varela
Elena Varlinskaya
Leigh Ann Vaughn
René Veenstra
Beth Venzke
Beatrix Vereijken
Lynne Vernon-Feagans
Marilyn Vihman
Peter Vishton
Frank Vitaro
Fred Vondracek
Robert Vorona
Athena Vouloumanos
Cecilia Wainryb
Laurie Wakschlag
Jane Waldfogel
Arlene Walker-Andrews
Kim Wallen
Sharon Walpole
Jian Wang
John Wann
Dorothy Warner
Daniel Waschbusch
Julie Washington
Gail Wasserman
Sarah Watamura
Amanda Waterman
Harriet Waters
Norman Watt
Niobe Way
Farrell Webb
Sara Webb
Catherine Weir
Thomas Weisner
Daniel Weiss
Marilyn Welsh
Jerry West
Barbara White
Leanne Whiteside-Mansell
Keith Widaman
Teresa Wilcox
Justin Williams
Beverly Wilson
Mark Wilson
Melvin Wilson
Jennifer Windsor
Ellen Winner
James Winslow
Joel Winston
Melissa Witkow
Dieter Wolke
Susan Woodhouse
Jacqueline Woolley
Margaret Wright
Lawrence Wu
Karen Wynn
Jenny Yau
Wei-Jun Yeung
Tiffany Yip
Paul Yoder
Albert Yonas
Hirokazu Yoshikawa
Liane Young
Lise Youngblade
Galit Yovel
Charles Zeanah
Debra Zeifman
Janice Zeman
Qing Zhou
Yair Ziv

Index to Volume 79

Abbott, Robert D. *See* Gruman, Diana H.
Abbot-Smith, Kirsten *See* Dittmar, Miriam
Aber, J. Lawrence *See* Clements, Margaret
Abrams, Dominic; Rutland, Adam; Ferrell, Jennifer M.; & Pelletier, Joseph. Children's Judgments of Disloyal and Immoral Peer Behavior: Subjective Group Dynamics in Minimal Intergroup Contexts, 444–461
Adams, Erin E. *See* Kochanska, Grazyna
Adolph, Karen E.; & Robinson, Scott R. In Defense of Change Processes, 1648–1653
Aikins, Julie Wargo *See* Simon, Valerie A.
Aksan, Nazan *See* Kochanska, Grazyna
Allen, Joseph P. *See* McElhaney, Kathleen B.
Allen, Nicholas B. *See* Yap, Marie B. H.
Anderson, Daniel R. *See* Schmidt, Marie Evans
Anderson, David I. *See* Campos, Joseph J
Angelini, Laura *See* Ghetti, Simona
Antonishak, Jill *See* McElhaney, Kathleen B.
Astington, Janet Wilde *See* Filippova, Eva
Atzaba-Poria, Naama; & Pike, Alison. Correlates of Parental Differential Treatment: Parental and Contextual Factors During Middle Childhood, 217–232

Bakermans-Kranenburg, Marian J. *See* Beijersbergen, Mariëlle D.
Baldelomar, Oscar *See* Fuligni, Andrew J.
Balestrino, Maria D. *See* Loe, Irene M.
Bao, Xue-hua; & Lam, Shui-fong. Who Makes the Choice? Rethinking the Role of Autonomy and Relatedness in Chinese Children's Motivation, 269–283
Barbarin, Oscar A. *See* Mashburn, Andrew J.
Barbu-Roth, Marianne *See* Campos, Joseph J
Barcelo, Hélène *See* Hanish, Laura D.
Barker, Edward D. *See* Joussemet, Mireille
Barner, David; & Snedeker, Jesse. Compositionality and Statistics in Adjective Acquisition: 4-Year-Olds Interpret *Tall* and *Short* Based on the Size Distributions of Novel Noun Referents, 594–608
Barner, Stacy *See* Snyder, James
Bates, John E. *See* Davis-Kean, Pamela E.
Bates, John E. *See* Fontaine, Reid Griffith
Bauman, Karl E. *See* Ennett, Susan T.
Beijersbergen, Mariëlle D.; Bakermans-Kranenburg, Marian J.; van IJzendoorn, Marinus H.; & Juffer, Femmie. Stress Regulation in Adolescents: Physiological Reactivity During the Adult Attachment Interview and Conflict Interaction, 1707–1720
Bergeron, Caroline *See* Murray, Lynne
Betjemann, Rebecca S.; & Keenan, Janice M. Phonological and Semantic Priming in Children With Reading Disability, 1086–1102
Bierman, Karen L.; Domitrovich, Celene E.; Nix, Robert L.; Gest, Scott D.; Welsh, Janet A.; Greenberg, Mark T.; Blair, Clancy; Nelson, Keith E.; & Gill, Sukhdeep. Promoting Academic and Social-Emotional School Readiness: The Head Start REDI Program, 1802–1817
Blair, Clancy *See* Bierman, Karen L.
Boivin, Michel *See* Brendgen, Mara
Booth, Julie L.; & Siegler, Robert S. Numerical Magnitude Representations Influence Arithmetic Learning, 1016–1031
Booth-LaForce, Cathryn *See* McElwain, Nancy L.
Brannon, Elizabeth M. *See* Cordes, Sara
Brendgen, Mara; Boivin, Michel; Vitaro, Frank; Bukowski, William M.; Dionne, Ginette; Tremblay, Richard E.; & Pérusse, Daniel. Linkages Between Children's and Their Friends' Social and Physical Aggression: Evidence for a Gene–Environment Interaction?, 13–29
Brown, B. Bradford; Herman, Melissa; Hamm, Jill V.; & Heck, Daniel J. Ethnicity and Image: Correlates of Crowd Affiliation Among Ethnic Minority Youth, 529–546
Brown, Christia Spears *See* Leaper, Campbell
Bryant, Donna *See* Mashburn, Andrew J.
Bukowski, William M. *See* Brendgen, Mara
Burchinal, Margaret *See* Mashburn, Andrew J.
Burt, Keith B.; Obradović, Jelena; Long, Jeffrey D.; & Masten, Ann S. The Interplay of Social Competence and Psychopathology Over 20 Years: Testing Transactional and Cascade Models, 359–374
Bushman, Brad J. *See* Thomaes, Sander
Butler, Ruth; & Shalit-Naggar, Rachel. Gender and Patterns of Concerned Responsiveness in Representations of the Mother–Daughter and Mother–Son Relationship, 836–851
Buttelmann, David; Carpenter, Malinda; Call, Josep; & Tomasello, Michael. Rational Tool Use and Tool Choice in Human Infants and Great Apes, 609–626

Caeyenberghs, Karen *See* van Roon, Dominique
Cai, Li *See* Ennett, Susan T.
Calkins, Susan D. *See* Degnan, Kathryn A.
Calkins, Susan D. *See* Propper, Cathi
Call, Josep *See* Buttelmann, David
Campos, Joseph J.; Witherington, David; Anderson, David I; Frankel, Carl I; Uchiyama, Ichiro; & Barbu-Roth, Marianne. Rediscovering Development in Infancy, 1625–1632
Caprara, Gian Vittorio *See* Paciello, Marinella
Carbone, Mary Anna *See* Propper, Cathi
Carpenter, Malinda *See* Buttelmann, David
Card, Noel A.; Stucky, Brian D.; Sawalani, Gita M.; & Little, Todd D. Direct and Indirect Aggression During Childhood and Adolescence: A Meta-Analytic Review of Gender Differences, Intercorrelations, and Relations to Maladjustment, 1185–1229
Cassidy, Adam R. *See* Cullerton-Sen, Crystal
Catalano, Richard F. *See* Gruman, Diana H.
Chang, Lei *See* Chen, Xinyin
Chaves-Gnecco, Diego *See* Loe, Irene M.
Chen, Xinyin; Chang, Lei; Liu, Hongyun; & He, Yunfeng. Effects of the Peer Group on the Development of Social Functioning and Academic Achievement: A Longitudinal Study in Chinese Children, 235–251
Cicchetti, Dante *See* Cullerton-Sen, Crystal

Cicchetti, Dante *See* Davies, Patrick T.
Cicchetti, Dante *See* Valentino, Kristin
Clark, Caron A. C.; Woodward, Lianne J.; Horwood, L. John; & Moor, Stephanie. Development of Emotional and Behavioral Regulation in Children Born Extremely Preterm and Very Preterm: Biological and Social Influences, 1444–1462
Clements, Margaret; Aber, J. Lawrence; & Seidman, Edward. The Dynamics of Life Stressors and Depressive Symptoms in Early Adolescence: A Test of Six Theoretical Models, 1168–1182
Collins, W. Andrew *See* Davis-Kean, Pamela E.
Conley, Dalton *See* Yeung, W. Jean
Connell, Arin *See* Dishion, Thomas J.
Connolly, Jennifer *See* Pepler, Debra
Converse, Alexander K. *See* Schneider, Mary L.
Cooper, Peter J. *See* Murray, Lynne
Coplan, Robert J. *See* Hastings, Paul D.
Corbett, Michael *See* Morrongiello, Barbara A.
Cordes, Sara; & Brannon, Elizabeth M. The Difficulties of Representing Continuous Extent in Infancy: Using Number Is Just Easier, 476–489
Corrigall, Kathleen *See* Hall, D. Geoffrey
Côté, Sylvana *See* Joussemet, Mireille
Cox, Martha *See* Propper, Cathi
Craig, Wendy *See* Pepler, Debra
Crick, Nicki R. *See* Cullerton-Sen, Crystal
Crosnoe, Robert; Riegle-Crumb, Catherine; Field, Sam; Frank, Kenneth; & Muller, Chandra. Peer Group Contexts of Girls' and Boys' Academic Experiences, 139 –155
Croudace, Tim *See* Golombok, Susan
Cullerton-Sen, Crystal; Cassidy, Adam R.; Murray-Close, Dianna; Cicchetti, Dante; Crick, Nicki R.; & Rogosch, Fred A. Childhood Maltreatment and the Development of Relational and Physical Aggression: The Importance of a Gender-Informed Approach, 1736–1751
Cummings, E. Mark *See* Davies, Patrick T.
Cumsille, Patricio *See* Darling, Nancy

Dahinten, V. Susan *See* Kohen, Dafna E.
Daniel, Ella *See* Knafo, Ariel
Darling, Nancy; Cumsille, Patricio; & Martínez, M. Loreto. Individual Differences in Adolescents' Beliefs About the Legitimacy of Parental Authority and Their Own Obligation to Obey: A Longitudinal Investigation, 1103–1118
David, Barbara J. *See* Grace, Diana M.
Davies, Patrick T.; Sturge-Apple, Melissa L.; Cicchetti, Dante; & Cummings, E. Mark. Adrenocortical Underpinnings of Children's Psychological Reactivity to Interparental Conflict, 1693–1706
Davies, Patrick T.; Woitach, Meredith J.; Winter, Marcia A.; & Cummings, E. Mark. Children's Insecure Representations of the Interparental Relationship and Their School Adjustment: The Mediating Role of Attention Difficulties, 1570–1582
Davis-Kean, Pamela E.; Huesmann, L. Rowell; Jager, Justin; Collins, W. Andrew; Bates, John E.; & Lansford, Jennifer E. Changes in the Relation of Self-Efficacy Beliefs and Behaviors Across Development, 1257–1269
Degnan, Kathryn A.; Calkins, Susan D.; Keane, Susan P.; & Hill-Soderlund, Ashley L. Profiles of Disruptive Behavior Across Early Childhood: Contributions of Frustration Reactivity, Physiological Regulation, and Maternal Behavior, 1357–1376
DeJesus, Onofre T. *See* Schneider, Mary L.
Deng, Xianli *See* Zhou, Qing
de Rosnay, Marc *See* Murray, Lynne
Dionne, Ginette *See* Brendgen, Mara
Dishion, Thomas J.; Shaw, Daniel; Connell, Arin; Gardner, Frances; Weaver, Chelsea; & Wilson, Melvin. The Family Check-Up With High-Risk Indigent Families: Preventing Problem Behavior by Increasing Parents' Positive Behavior Support in Early Childhood, 1395–1414
Dittmar, Miriam; Abbot-Smith, Kirsten; Lieven, Elena; & Tomasello, Michael. German Children's Comprehension of Word Order and Case Marking in Causative Sentences, 1152–1167
Dodge, Kenneth A. *See* Fontaine, Reid Griffith
Dodge, Kenneth A.; Greenberg, Mark T.; Malone, Patrick S. & Conduct Problems Prevention Research Group. Testing an Idealized Dynamic Cascade Model of the Development of Serious Violence in Adolescence, 1907–1927
Domitrovich, Celene E. *See* Bierman, Karen L.
Donegan, Eleanor *See* Hall, D. Geoffrey
Downer, Jason T. *See* Mashburn, Andrew J.
Dupéré, Véronique; Lacourse, Éric; Willms, J. Douglas; Leventhal, Tama; & Tremblay, Richard E. Neighborhood Poverty and Early Transition to Sexual Activity in Young Adolescents: A Developmental Ecological Approach, 1463–1476
DuRant, Robert *See* Ennett, Susan T.

Early, Diane M. *See* Mashburn, Andrew J.
Eisenberg, Nancy *See* Zhou, Qing
Ennett, Susan T.; Foshee, Vangie A.; Bauman, Karl E.; Hussong, Andrea; Faris, Robert; Hipp, John; Cai, Li; & McNaughton Reyes, Heathe Luz. DuRant, Robert; The Social Ecology of Adolescent Alcohol Misuse, 1777–1791
Ensor, Rosie; & Hughes, Claire. Content or Connectedness? Mother–Child Talk and Early Social Understanding, 201–216

Fabes, Richard A. *See* Hanish, Laura D.
Faris, Robert *See* Ennett, Susan T.
Feldman, Heidi M. *See* Loe, Irene M.
Ferrell, Jennifer M. *See* Abrams, Dominic
Fida, Roberta *See* Paciello, Marinella
Field, Sam *See* Crosnoe, Robert
Filippova, Eva; & Astington, Janet Wilde. Further Development in Social Reasoning Revealed in Discourse Irony Understanding, 126–138
Fisher, Anna V. *See* Sloutsky, Vladimir M.
Fleming, Charles B. *See* Gruman, Diana H.
Flook, Lisa; & Fuligni, Andrew J. Family and School Spillover in Adolescents' Daily Lives, 776–787
Fontaine, Reid Griffith; Yang, Chongming; Dodge, Kenneth A.; Bates, John E.; & Pettit, Gregory S. Testing an Individual Systems Model of Response Evaluation and Decision (RED) and Antisocial Behavior Across Adolescence, 462–475
Foshee, Vangie A. *See* Ennett, Susan T.
Frank, Kenneth *See* Crosnoe, Robert
Frankel, Carl I *See* Campos, Joseph J.

Frye, Douglas *See* Ziv, Margalit
Fuligni, Andrew J. *See* Flook, Lisa
Fuligni, Andrew J.; Kiang, Lisa; Witkow, Melissa R.; & Baldelomar, Oscar. Stability and Change in Ethnic Labeling Among Adolescents From Asian and Latin American Immigrant Families, 944–956

Gajewski, Lisa L. *See* Schneider, Mary L.
Gallagher, Kathleen Cranley *See* Stright, Anne Dopkins
Gallop, Robert *See* Maikovich, Andrea Kohn
Galloway, James C *See* Lobo, Michele A.
Gardner, Frances *See* Dishion, Thomas J.
Gauvain, Mary; & Perez, Susan M. Mother–Child Planning and Child Compliance, 761–775
Gelman, Susan A. *See* Rhodes, Marjorie
Gest, Scott D. *See* Bierman, Karen L.
Ghetti, Simona; & Angelini, Laura. The Development of Recollection and Familiarity in Childhood and Adolescence: Evidence From the Dual-Process Signal Detection Model, 339–358
Giesbrecht, Gerald *See* Müller, Ulrich
Gill, Sukhdeep *See* Bierman, Karen L.
Goh, Wendy *See* Kuhn, Deanna
Golding, Jean *See* Golombok, Susan
Golinkoff, Roberta Michnick *See* Imai, Mutsumi
Golombok, Susan; Rust, John; Zervoulis, Karyofyllis; Croudace, Tim; Golding, Jean; & Hines, Melissa. Developmental Trajectories of Sex-Typed Behavior in Boys and Girls: A Longitudinal General Population Study of Children Aged 2.5–8 Years, 1583–1593
Grace, Diana M.; David, Barbara J.; & Ryan, Michelle K. Investigating Preschoolers' Categorical Thinking About Gender Through Imitation, Attention, and the Use of Self-Categories, 1928–1941
Granger, Douglas A. *See* Lisonbee, Jared A.
Greenberg, Mark T. *See* Bierman, Karen L.
Greenberg, Mark T. *See* Dodge, Kenneth A.
Gruman, Diana H.; Harachi, Tracy W.; Abbott, Robert D.; Catalano, Richard F.; & Fleming, Charles B. Longitudinal Effects of Student Mobility on Three Dimensions of Elementary School Engagement, 1833–1852
Gummerum, Michaela; Keller, Monika; Takezawa, Masanori; & Mata, Jutta. To Give or Not to Give: Children's and Adolescents' Sharing and Moral Negotiations in Economic Decision Situations, 562–576

Hall, D. Geoffrey; Corrigall, Kathleen; Rhemtulla, Mijke; Donegan, Eleanor; & Xu, Fei. Infants' Use of Lexical-Category-to-Meaning Links in Object Individuation, 1432–1443
Halpern, Carolyn Tucker *See* Propper, Cathi
Hamm, Jill V. *See* Brown, B. Bradford
Hamre, Bridget K. *See* Mashburn, Andrew J.
Hanish, Laura D.; Martin, Carol Lynn; Fabes, Richard A.; & Barcelo, Hélène. The Breadth of Peer Relationships Among Preschoolers: An Application of the *Q*-Connectivity Method to Externalizing Behavior, 1119–1136
Harachi, Tracy W. *See* Gruman, Diana H.
Haryu, Etsuko *See* Imai, Mutsumi
Hastings, Paul D.; Sullivan, Caroline; McShane, Kelly E.; Coplan, Robert J.; Utendale, William T.; & Vyncke, Johanna D. Parental Socialization, Vagal Regulation, and Preschoolers' Anxious Difficulties: Direct Mothers and Moderated Fathers, 45–64
Hayne, Harlene *See* Reese, Elaine
He, Yunfeng *See* Chen, Xinyin
Heck, Daniel J. *See* Brown, B. Bradford
Herald-Brown, Sarah L. *See* Ladd, Gary W.
Herman, Melissa *See* Brown, B. Bradford
Hill-Soderlund, Ashley L. *See* Degnan, Kathryn A.
Hill-Soderlund, Ashley L. *See* Propper, Cathi
Hines, Melissa *See* Golombok, Susan
Hipp, John *See* Ennett, Susan T.
Hirsh-Pasek, Kathy *See* Imai, Mutsumi
Hodges, Ernest V. E. *See* Peets, Kätlin
Hoefnagels, Cees *See* van Doesum, Karin T. M.
Hoehl, Stefanie; & Striano, Tricia. Neural Processing of Eye Gaze and Threat-Related Emotional Facial Expressions in Infancy, 1752–1760
Hooppell, Catherine *See* Schulz, Laura E.
Horwood, L. John *See* Clark, Caron A. C.
Hosman, Clemens M. H. *See* van Doesum, Karin T. M.
Howe, Mark L. Visual Distinctiveness and the Development of Children's False Memories, 65–79
Howes, Carollee *See* Mashburn, Andrew J.
Huesmann, L. Rowell *See* Davis-Kean, Pamela E.
Hughes, Claire *See* Ensor, Rosie
Hussong, Andrea *See* Ennett, Susan T.
Huston, Aletha C. From Research to Policy and Back, 1–12

Imai, Mutsumi; Li Lianjing; Haryu, Etsuko; Okada, Hiroyuki; Hirsh-Pasek, Kathy; Golinkoff, Roberta Michnick; & Shigematsu, Jun. Novel Noun and Verb Learning in Chinese-, English-, and Japanese-Speaking Children, 979–1000
Iordanou, Kalypso *See* Kuhn, Deanna

Jacobson, Joseph L. *See* Yumoto, Chie
Jacobson, Sandra W. *See* Yumoto, Chie
Jaffee, Sara R. *See* Maikovich, Andrea Kohn
Jager, Justin *See* Davis-Kean, Pamela E.
Jaswal, Vikram K.; McKercher, David A.; & VanderBorght, Mieke. Limitations on Reliability: Regularity Rules in the English Plural and Past Tense, 750–760
Jenkins, Adrianna C. *See* Schulz, Laura E.
Jiang, Depeng *See* Pepler, Debra
Johnson, Kassy *See* Snyder, James
Johnson, Scott P. *See* Soska, Kasey C.
Joussemet, Mireille; Vitaro, Frank; Barker, Edward D.; Côté, Sylvana; Nagin, Daniel S.; Zoccolillo, Mark; & Tremblay, Richard E. Controlling Parenting and Physical Aggression During Elementary School, 411–425
Juffer, Femmie *See* Beijersbergen, Mariëlle D.
Justin Dyer, W. *See* McElwain, Nancy L.

Kagan, Jerome. In Defense of Qualitative Changes in Development, 1606–1624
Kalish, Charles W.; & Lawson, Christopher A. Development of Social Category Representations: Early Appreciation of Roles and Deontic Relations, 577–593
Keane, Susan P. *See* Degnan, Kathryn A.
Keil, Frank C. *See* Newman, George E.
Keenan, Janice M. *See* Betjemann, Rebecca S.
Keller, Monika *See* Gummerum, Michaela
Kelley, Ken *See* Stright, Anne Dopkins
Khoury-Kassabri, Mona *See* Knafo, Ariel
Kiang, Lisa *See* Fuligni, Andrew J.
Kirkorian, Heather L. *See* Schmidt, Marie Evans

Klemencic, Nora *See* Morrongiello, Barbara A.
Knafo, Ariel; Daniel, Ella; & Khoury-Kassabri, Mona. Values as Protective Factors Against Violent Behavior in Jewish and Arab High Schools in Israel, 652–667
Kochanska, Grazyna; Aksan, Nazan; Prisco, Theresa R.; & Adams, Erin E. Mother–Child and Father–Child Mutually Responsive Orientation in the First 2 Years and Children's Outcomes at Preschool Age: Mechanisms of Influence, 30–44
Kohen, Dafna E.; Leventhal, Tama; Dahinten, V. Susan; & McIntosh, Cameron N. Neighborhood Disadvantage: Pathways of Effects for Young Children, 156–169
Kuhn, Deanna; Goh, Wendy; Iordanou, Kalypso; & Shaenfield, David. Arguing´ on the Computer: A Microgenetic Study of Developing Argument Skills in a Computer-Supported Environment, 1310–1328
Kurs-Lasky, Marcia *See* Loe, Irene M.

Lacourse, Éric *See* Dupéré, Véronique
Ladd, Gary W.; Herald-Brown, Sarah L.; & Reiser, Mark. Does Chronic Classroom Peer Rejection Predict the Development of Children's Classroom Participation During the Grade School Years?, 1001–1015
Ladouceur, Cecile D. *See* Yap, Marie B. H.
Lagattuta, Kristin Hansen *See* Sayfan, Liat
Laible, Deborah; Panfile, Tia; & Makariev, Drika. The Quality and Frequency of Mother–Toddler Conflict: Links With Attachment and Temperament, 426–443
Lam, Shui-fong *See* Bao, Xue-hua
Lam, Virginia L. *See* Leman, Patrick J.
Lansford, Jennifer E. *See* Davis-Kean, Pamela E.
Lansford, Jennifer E. *See* McElwain, Nancy L.
Larson, Julie A. *See* Schneider, Mary L.
Lawson, Christopher A. *See* Kalish, Charles W.
Leaper, Campbell; & Brown, Christia Spears. Perceived Experiences With Sexism Among Adolescent Girls, 685–704
Lee, Kang *See* Talwar, Victoria
Leman, Patrick J.; & Lam, Virginia L. The Influence of Race and Gender on Children's Conversations and Playmate Choices, 1329–1343
Leo, Irene *See* Macchi Cassia, Viola
Leventhal, Tama *See* Dupéré, Véronique
Leventhal, Tama *See* Kohen, Dafna E.
Lewis, Charlie *See* Oh, Seungmi
Lewkowicz, David J. Perception of Dynamic and Static Audiovisual Sequences in 3- and 4-Month-Old Infants, 1538–1554
Li-Grining, Christine P. *See* Votruba-Drzal, Elizabeth
Li Lianjing, *See* Imai, Mutsumi
Liben, Lynn S. *See* Myers, Lauren J.
Liben, Lynn S. Continuities and Discontinuities in Children and Scholarship, 1600–1605
Liben, Lynn S. Reflections on Child Development: The Journal and The Field, 1597–1599
Lieven, Elena *See* Dittmar, Miriam
Lisonbee, Jared A.; Mize, Jacquelyn; Payne, Amie Lapp; & Granger, Douglas A. Children's Cortisol and the Quality of Teacher–Child Relationships in Child Care, 1818–1832
Little, Todd D. *See* Card, Noel A.
Liu, Hongyun *See* Chen, Xinyin
Lobo, Michele A.; & Galloway, James C. Postural and Object-Oriented Experiences Advance Early Reaching, Object Exploration, and Means – End Behavior, 1869–1890
Loe, Irene M.; Balestrino, Maria D.; Phelps, Randall A.; Kurs-Lasky, Marcia; Chaves-Gnecco, Diego; Paradise, Jack L.; & Feldman, Heidi M. Early Histories of School-Aged Children With Attention-Deficit/Hyperactivity Disorder, 1853–1868
Long, Jeffrey D. *See* Burt, Keith B.
Lugo-Gil, Julieta; & Tamis-LeMonda, Catherine S. Family Resources and Parenting Quality: Links to Children's Cognitive Development Across the First 3 Years, 1065–1085
Lund, Anne Frankenfield *See* Schmidt, Marie Evans
Lupinetti, Catia *See* Paciello, Marinella
Lyon, Thomas D.; Malloy, Lindsay C.; Quas, Jodi A.; & Talwar, Victoria A. Coaching, Truth Induction, and Young Maltreated Children's False Allegations and False Denials, 914–929

Macchi Cassia, Viola; Valenza, Eloisa; Simion, Francesca; & Leo, Irene. Congruency as a Nonspecific Perceptual Property Contributing to Newborns' Face Preference, 807–820
MacDonald, Shelley *See* Reese, Elaine
Maikovich, Andrea Kohn; Jaffee, Sara R.; Odgers, Candice L.; & Gallop, Robert. Effects of Family Violence on Psychopathology Symptoms in Children Previously Exposed to Maltreatment, 1498–1512
Makariev, Drika *See* Laible, Deborah
Maldonado-Carreño, Carolina *See* Votruba-Drzal, Elizabeth
Malloy, Lindsay C. *See* Lyon, Thomas D.
Malone, Patrick S *See* Dodge, Kenneth A
Mark Cummings, E. *See* Davies, Patrick T.
Martin, Carol Lynn *See* Hanish, Laura D.
Martínez, M. Loreto *See* Darling, Nancy
Mashburn, Andrew J.; Pianta, Robert C.; Hamre, Bridget K.; Downer, Jason T.; Barbarin, Oscar A.; Bryant, Donna; Burchinal, Margaret; Early, Diane M.; & Howes, Carollee. Measures of Classroom Quality in Prekindergarten and Children's Development of Academic, Language, and Social Skills, 732–749
Masnick, Amy M.; & Morris, Bradley J. Investigating the Development of Data Evaluation: The Role of Data Characteristics, 1032–1048
Masten, Ann S. *See* Burt, Keith B.
Mata, Jutta *See* Gummerum, Michaela
McColgan, Kerry L.; & McCormack, Teresa. Searching and Planning: Young Children's Reasoning About Past and Future Event Sequences, 1477–1497
McCormack, Teresa *See* McColgan, Kerry L.
McEachern, Amber *See* Snyder, James
McElhaney, Kathleen B.; Antonishak, Jill; & Allen, Joseph P. "They Like Me, They Like Me Not": Popularity and Adolescents' Perceptions of Acceptance Predicting Social Functioning Over Time, 720–731
McElwain, Nancy L.; Booth-LaForce, Cathryn; Lansford, Jennifer E.; Wu, Xiaoying; & Dyer W. Justin. A Process Model of Attachment – Friend Linkages: Hostile Attribution Biases, Language Ability, and Mother – Child Affective Mutuality as Intervening Mechanisms, 1891–1906
McIntosh, Cameron N. *See* Kohen, Dafna E.
McKercher, David A. *See* Jaswal, Vikram K.
McNaughton Reyes, Heathe Luz *See* Ennett, Susan T.

McNeil, Nicole M. Limitations to Teaching Children 2 + 2 = 4: Typical Arithmetic Problems Can Hinder Learning of Mathematical Equivalence, 1524–1537
McShane, Kelly E. *See* Hastings, Paul D.
Miles, Sarah *See* Stipek, Deborah
Mills-Koonce, W. Roger *See* Propper, Cathi
Mize, Jacquelyn *See* Lisonbee, Jared A.
Mondloch, Catherine J.; & Thomson, Kendra. Limitations in 4-Year-Old Children's Sensitivity to the Spacing Among Facial Features, 1513–1523
Moore, Colleen F. *See* Schneider, Mary L.
Moor, Stephanie *See* Clark, Caron A. C.
Moore, Ginger A. *See* Propper, Cathi
Morris, Bradley J. *See* Masnick, Amy M.
Morrongiello, Barbara A.; Klemencic, Nora; & Corbett, Michael. Interactions Between Child Behavior Patterns and Parent Supervision: Implications for Children's Risk of Unintentional Injury, 627–638
Muller, Chandra *See* Crosnoe, Robert
Müller, Ulrich; & Giesbrecht, Gerald. Methodological and Epistemological Issues in the Interpretation of Infant Cognitive Development, 1654–1658
Murray, Lynne; de Rosnay, Marc; Pearson, Joanna; Bergeron, Caroline; Schofield, Elizabeth; Royal-Lawson, Melanie; & Cooper, Peter J. Intergenerational Transmission of Social Anxiety: The Role of Social Referencing Processes in Infancy, 1049–1064
Murray-Close, Dianna *See* Cullerton-Sen, Crystal
Myers, Lauren J.; & Liben, Lynn S. The Role of Intentionality and Iconicity in Children's Developing Comprehension and Production of Cartographic Symbols, 668–684

Nagin, Daniel S. *See* Joussemet, Mireille
Nelson, Keith E. *See* Bierman, Karen L.
Newman, George E.; & Keil, Frank C. Where Is the Essence? Developmental Shifts in Children's Beliefs About Internal Features, 1344–1356
Nix, Robert L. *See* Bierman, Karen L.
Nummenmaa, Lauri; Peets, Kätlin; & Salmivalli, Christina. Automatic Activation of Adolescents' Peer-Relational Schemas: Evidence From Priming With Facial Identity, 1659–1675

Obradović, Jelena *See* Burt, Keith B.
Odgers, Candice L. *See* Maikovich, Andrea Kohn
Oh, Seungmi; & Lewis, Charlie. Korean Preschoolers' Advanced Inhibitory Control and Its Relation to Other Executive Skills and Mental State Understanding, 80–99
Okada, Hiroyuki *See* Imai, Mutsumi
Olthof, Tjeert *See* Thomaes, Sander
Opfer, John E.; & Thompson, Clarissa A. The Trouble With Transfer: Insights From Microgenetic Changes in the Representation of Numerical Magnitude, 788–804
Ouellette, Gene; & Sénéchal, Monique. Pathways to Literacy: A Study of Invented Spelling and Its Role in Learning to Read, 899–913

Paciello, Marinella; Fida, Roberta; Tramontano, Carlo; Lupinetti, Catia; & Caprara, Gian Vittorio. Stability and Change of Moral Disengagement and Its Impact on Aggression and Violence in Late Adolescence, 1288–1309
Panfile, Tia *See* Laible, Deborah
Paradise, Jack L. *See* Loe, Irene M.
Payne, Amie Lapp *See* Lisonbee, Jared A.
Pearson, Joanna *See* Murray, Lynne
Peets, Kätlin *See* Nummenmaa, Lauri
Peets, Kätlin; Hodges, Ernest V. E.; & Salmivalli, Christina. Affect-Congruent Social-Cognitive Evaluations and Behaviors, 170–185
Pelletier, Joseph *See* Abrams, Dominic
Pempek, Tiffany A. *See* Schmidt, Marie Evans
Pepler, Debra; Jiang, Depeng; Craig, Wendy; & Connolly, Jennifer. Developmental Trajectories of Bullying and Associated Factors, 325–338
Perez, Susan M. *See* Gauvain, Mary
Perone, Sammy *See* Spencer, John P.
Pérusse, Daniel *See* Brendgen, Mara
Pettit, Gregory S. *See* Fontaine, Reid Griffith
Phelps, Randall A. *See* Loe, Irene M.
Pianta, Robert C. *See* Mashburn, Andrew J.
Pike, Alison *See* Atzaba-Poria, Naama
Posada, Roberto; & Wainryb, Cecilia. Moral Development in a Violent Society: Colombian Children's Judgments in the Context of Survival and Revenge, 882–898
Prinstein, Mitchell J. *See* Simon, Valerie A.
Prisco, Theresa R. *See* Kochanska, Grazyna
Propper, Cathi; Moore, Ginger A.; Mills-Koonce, W. Roger; Halpern, Carolyn Tucker; Hill-Soderlund, Ashley L.; Calkins, Susan D.; Carbone, Mary Anna; & Cox, Martha. Gene–Environment Contributions to the Development of Infant Vagal Reactivity: The Interaction of Dopamine and Maternal Sensitivity, 1377–1394
Provines, Jessica *See* Snyder, James

Quas, Jodi A. *See* Lyon, Thomas D.
Quinn, Paul C. In Defense of Core Competencies, Quantitative Change, and Continuity, 1633–1638

Ramani, Geetha B.; & Siegler, Robert S. Promoting Broad and Stable Improvements in Low-Income Children's Numerical Knowledge Through Playing Number Board Games, 375–394
Reese, Elaine; Hayne, Harlene; & MacDonald, Shelley. Looking Back to the Future: Māori and Pakeha Mother–Child Birth Stories, 114–125
Reiser, Mark *See* Ladd, Gary W.
Rhemtulla, Mijke *See* Hall, D. Geoffrey
Rhoades, Kimberly A. Children's Responses to Interparental Conflict: A Meta-Analysis of Their Associations With Child Adjustment, 1942–1956
Rhodes, Marjorie; & Gelman, Susan A. Categories Influence Predictions About Individual Consistency, 1270–1287
Riegle-Crumb, Catherine *See* Crosnoe, Robert
Riksen-Walraven, J. Marianne *See* van Doesum, Karin T. M.
Roberts, Andrew D. *See* Schneider, Mary L.
Robinson, Scott R. *See* Adolph, Karen E.
Rogosch, Fred A. *See* Cullerton-Sen, Crystal
Rogosch, Fred A. *See* Valentino, Kristin
Royal-Lawson, Melanie *See* Murray, Lynne
Ruffman, Ted *See* Taumoepeau, Mele
Rust, John *See* Golombok, Susan
Rutland, Adam *See* Abrams, Dominic
Ryan, Michelle K. *See* Grace, Diana M.

Salmivalli, Christina *See* Nummenmaa, Lauri
Salmivalli, Christina *See* Peets, Kätlin
Saudino, Kimberly J.; & Zapfe, Jeffrey A. Genetic Influences on Activity Level in Early Childhood: Do Situations Matter?, 930–943
Sayfan, Liat; & Lagattuta, Kristin Hansen. Grownups Are Not Afraid of Scary Stuff, but Kids Are: Young Children's and Adults' Reasoning About Children's, Infants', and Adults' Fears, 821–835
Sawalani, Gita M. *See* Card, Noel A.
Schmidt, Marie Evans; Pempek, Tiffany A.; Kirkorian, Heather L.; Lund, Anne Frankenfield; & Anderson, Daniel R. The Effects of Background Television on the Toy Play Behavior of Very Young Children, 1137–1151
Schneider, Mary L.; Moore, Colleen F.; Gajewski, Lisa L.; Larson, Julie A.; Roberts, Andrew D.; Converse, Alexander K.; & DeJesus, Onofre T. Sensory Processing Disorder in a Primate Model: Evidence From a Longitudinal Study of Prenatal Alcohol and Prenatal Stress Effects, 100–113
Schofield, Elizabeth *See* Murray, Lynne
Schrepferman, Lynn *See* Snyder, James
Schulz, Laura E.; Hooppell, Catherine; & Jenkins, Adrianna C. Judicious Imitation: Children Differentially Imitate Deterministically and Probabilistically Effective Actions, 395–410
Seidman, Edward *See* Clements, Margaret
Sénéchal, Monique *See* Ouellette, Gene
Shalit-Naggar, Rachel *See* Butler, Ruth
Shaenfield, David *See* Kuhn, Deanna
Shaw, Daniel *See* Dishion, Thomas J.
Shigematsu, Jun *See* Imai, Mutsumi
Siegler, Robert S. *See* Booth, Julie L.
Siegler, Robert S. *See* Ramani, Geetha B.
Simion, Francesca *See* Macchi Cassia, Viola
Simon, Valerie A.; Aikins, Julie Wargo; & Prinstein, Mitchell J. Romantic Partner Selection and Socialization During Early Adolescence, 1676–1692
Slooter, Luuk *See* Verkuyten, Maykel
Sloutsky, Vladimir M.; & Fisher, Anna V. Attentional Learning and Flexible Induction: How Mundane Mechanisms Give Rise to Smart Behaviors, 639–651
Smits-Engelsman, Bouwien C. M. *See* van Roon, Dominique
Snedeker, Jesse *See* Barner, David
Snyder, James; Schrepferman, Lynn; McEachern, Amber; Barner, Stacy; Johnson, Kassy; & Provines, Jessica. Peer Deviancy Training and Peer Coercion: Dual Processes Associated With Early-Onset Conduct Problems, 252–268
Solomon, Ayelet *See* Ziv, Margalit
Soska, Kasey C.; & Johnson, Scott P. Development of Three-Dimensional Object Completion in Infancy, 1230–1236
Spencer, John P.; & Perone, Sammy. Defending Qualitative Change: The View From Dynamical Systems Theory, 1639–1647
Spinath, Birgit; & Steinmayr, Ricarda. Longitudinal Analysis of Intrinsic Motivation and Competence Beliefs: Is There a Relation Over Time?, 1555–1569
Stegge, Hedy *See* Thomaes, Sander
Steinmayr, Ricarda *See* Spinath, Birgit
Stipek, Deborah; & Miles, Sarah. Effects of Aggression on Achievement: Does Conflict With the Teacher Make It Worse?, 1721–1735
Striano, Tricia *See* Hoehl, Stefanie
Stright, Anne Dopkins; Gallagher, Kathleen Cranley; & Kelley, Ken. Infant Temperament Moderates Relations Between Maternal Parenting in Early Childhood and Children's Adjustment in First Grade, 186–200
Stucky, Brian D. *See* Card, Noel A.
Sturge-Apple, Melissa L. *See* Davies, Patrick T.
Sullivan, Caroline *See* Hastings, Paul D.
Swinnen, Stephan P. *See* van Roon, Dominique

Takezawa, Masanori *See* Gummerum, Michaela
Talwar, Victoria A. *See* Lyon, Thomas D.
Talwar, Victoria; & Lee, Kang. Social and Cognitive Correlates of Children's Lying Behavior, 866–881
Tamis-LeMonda, Catherine S. *See* Lugo-Gil, Julieta
Taumoepeau, Mele; & Ruffman, Ted. Stepping Stones to Others' Minds: Maternal Talk Relates to Child Mental State Language and Emotion Understanding at 15, 24, and 33 Months, 284–302
Tein, Jenn-Yun *See* Zhou, Qing
Thomaes, Sander; Bushman, Brad J.; Stegge, Hedy; & Olthof, Tjeert. Trumping Shame by Blasts of Noise: Narcissism, Self-Esteem, Shame, and Aggression in Young Adolescents, 1792–1801
Thompson, Clarissa A. *See* Opfer, John E.
Thomson, Kendra *See* Mondloch, Catherine J.
Tomasello, Michael *See* Buttelmann, David
Tomasello, Michael *See* Dittmar, Miriam
Toth, Sheree L. *See* Valentino, Kristin
Tramontano, Carlo *See* Paciello, Marinella
Tremblay, Richard E. *See* Brendgen, Mara
Tremblay, Richard E. *See* Dupéré, Véronique
Tremblay, Richard E. *See* Joussemet, Mireille

Uchiyama, Ichiro *See* Campos, Joseph J
Utendale, William T. *See* Hastings, Paul D.

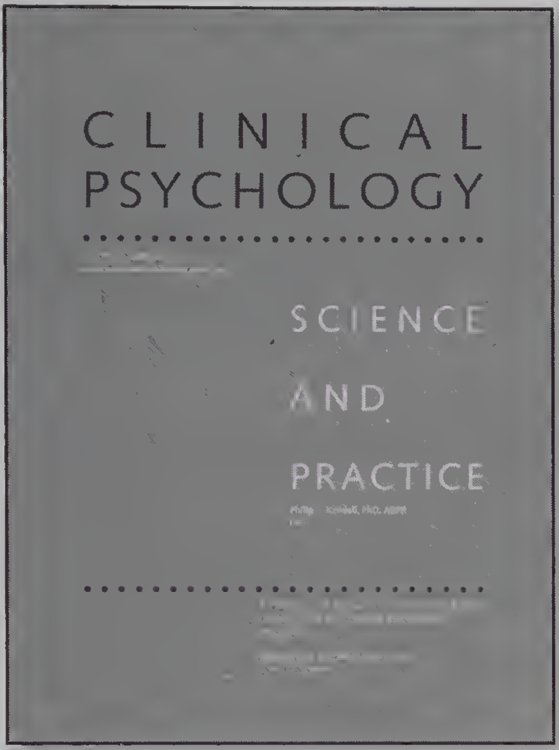
Valentino, Kristin; Cicchetti, Dante; Rogosch, Fred A.; & Toth, Sheree L. Memory, Maternal Representations, and Internalizing Symptomatology Among Abused, Neglected, and Nonmaltreated Children, 705–719
Valenza, Eloisa *See* Macchi Cassia, Viola
van Doesum, Karin T. M.; Riksen-Walraven, J. Marianne; Hosman, Clemens M. H.; & Hoefnagels, Cees. A Randomized Controlled Trial of a Home-Visiting Intervention Aimed at Preventing Relationship Problems in Depressed Mothers and Their Infants, 547–561
van IJzendoorn, Marinus H. *See* Beijersbergen, Mariëlle D.
van Roon, Dominique; Caeyenberghs, Karen; Swinnen, Stephan P.; & Smits-Engelsman, Bouwien C. M. Development of Feedforward Control in a Dynamic Manual Tracking Task, 852–865
VanderBorght, Mieke *See* Jaswal, Vikram K.
Verkuyten, Maykel; & Slooter, Luuk. Muslim and Non-Muslim Adolescents' Reasoning About Freedom of Speech and Minority Rights, 514–528
Vitaro, Frank *See* Brendgen, Mara
Vitaro, Frank *See* Joussemet, Mireille
Votruba-Drzal, Elizabeth; Li-Grining, Christine P.; & Maldonado-Carreño, Carolina. A Developmental Perspective on Full- Versus Part-Day Kindergarten and Children's Academic Trajectories Through Fifth Grade, 957–978
Vyncke, Johanna D. *See* Hastings, Paul D.

Wainryb, Cecilia *See* Posada, Roberto
Wang, Yun *See* Zhou, Qing
Weaver, Chelsea *See* Dishion, Thomas J.
Welsh, Janet A. *See* Bierman, Karen L.
Willms, J. Douglas *See* Dupéré, Véronique
Wilson, Melvin *See* Dishion, Thomas J.
Winter, Marcia A. *See* Davies, Patrick T.
Witherington, David *See* Campos, Joseph J.
Witkow, Melissa R. *See* Fuligni, Andrew J.
Woitach, Meredith J. See Davies, Patrick T.
Wolchik, Sharlene A. *See* Zhou, Qing
Woodward, Lianne J. *See* Clark, Caron A. C.
Wu, Xiaoying *See* McElwain, Nancy L.

Xu, Fei *See* Hall, D. Geoffrey

Yap, Marie B. H.; Allen, Nicholas B.; & Ladouceur, Cecile D. Maternal Socialization of Positive Affect: The Impact of Invalidation on Adolescent Emotion Regulation and Depressive Symptomatology, 1415–1431
Yang, Chongming *See* Fontaine, Reid Griffith
Yeung, W. Jean; & Conley, Dalton. Black–White Achievement Gap and Family Wealth, 303–324
Yumoto, Chie; Jacobson, Sandra W.; & Jacobson, Joseph L. Fetal Substance Exposure and Cumulative Environmental Risk in an African American Cohort, 1761–1776

Zapfe, Jeffrey A. *See* Saudino, Kimberly J.
Zervoulis, Karyofyllis *See* Golombok, Susan
Zhou, Qing; Wang, Yun; Deng, Xianli; Eisenberg, Nancy; Wolchik, Sharlene A.; & Tein, Jenn-Yun. Relations of Parenting and Temperament to Chinese Children's Experience of Negative Life Events, Coping Efficacy, and Externalizing Problems, 493–513
Ziv, Margalit; Solomon, Ayelet; & Frye, Douglas. Young Children's Recognition of the Intentionality of Teaching, 1237–1256
Zoccolillo, Mark *See* Joussemet, Mireille



CLINICAL PSYCHOLOGY

SCIENCE AND PRACTICE

CLINICAL PSYCHOLOGY:
SCIENCE AND PRACTICE

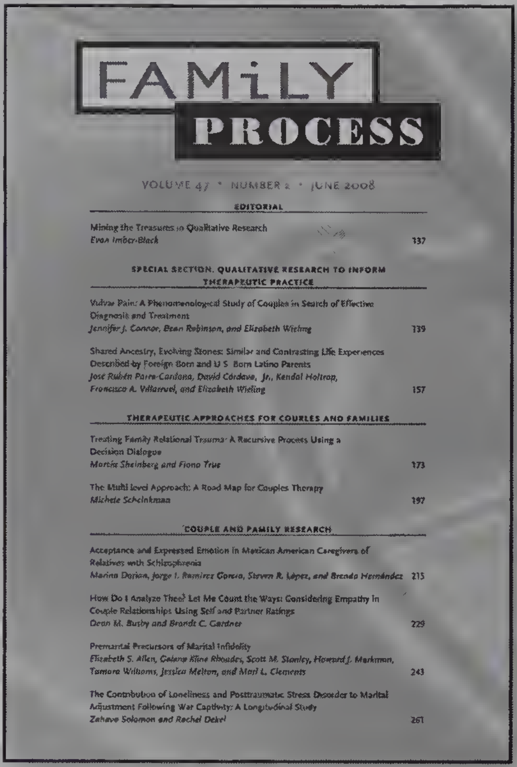
Edited by
PHILIP C. KENDALL

Published on behalf of the Society of Clinical Psychology, Division 12 of the American Psychological Association

Clinical Psychology: Science and Practice presents cutting-edge developments in the science and practice of clinical psychology by publishing scholarly topical reviews of research, theory, and application to diverse areas of the field, including assessment, intervention, service delivery, and professional issues.

WILEY-BLACKWELL

For more information and to subscribe online visit
www.blackwellpublishing.com/CPSP



FAMiLY
PROCESS

VOLUME 47 • NUMBER 2 • JUNE 2008

EDITORIAL

Mining the Treasures in Qualitative Research
Evan Imber-Black 137

SPECIAL SECTION. QUALITATIVE RESEARCH TO INFORM THERAPEUTIC PRACTICE

Vulvar Pain: A Phenomenological Study of Couples in Search of Effective Diagnosis and Treatment
Jennifer J. Connor, Bean Robinson, and Elizabeth Wieling 139

Shared Ancestry, Evolving Stories: Similar and Contrasting Life Experiences Described by Foreign Born and U.S. Born Latino Parents
José Rubén Parra-Cardona, David Córdova, Jr., Kendal Holtrop, Francisco A. Villarruel, and Elizabeth Wieling 157

THERAPEUTIC APPROACHES FOR COUPLES AND FAMILIES

Treating Family Relational Trauma: A Recursive Process Using a Decision Dialogue
Marcia Sheinberg and Fiona True 173

The Multi-level Approach: A Road Map for Couples Therapy
Michele Scheinkman 197

COUPLE AND FAMILY RESEARCH

Acceptance and Expressed Emotion in Mexican American Caregivers of Relatives with Schizophrenia
Marina Dorian, Jorge I. Ramírez García, Steven R. López, and Brenda Hernández 215

How Do I Analyze Thee? Let Me Count the Ways: Considering Empathy in Couple Relationships Using Self and Partner Ratings
Dean M. Busby and Brandt C. Gardner 229

Premarital Precursors of Marital Infidelity
Elizabeth S. Allen, Galena Kline Rhoades, Scott M. Stanley, Howard J. Markman, Tamara Williams, Jessica Melton, and Mari L. Clements 243

The Contribution of Loneliness and Posttraumatic Stress Disorder to Marital Adjustment Following War Captivity: A Longitudinal Study
Zahava Solomon and Rachel Dekel 261

FAMiLY
PROCESS

Edited by
EVAN IMBER-BLACK

Published on behalf of the Family Process Institute

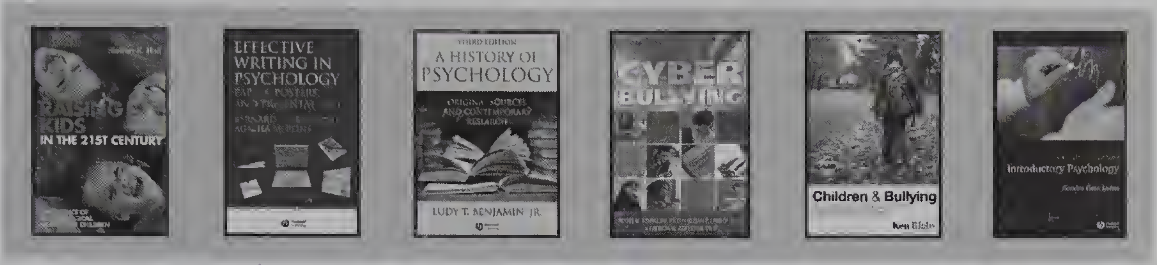
Family Process is an international, multidisciplinary, peer-reviewed journal committed to publishing original articles, including theory and practice, philosophical underpinnings, qualitative and quantitative clinical research, and training in couple and family therapy, family interaction, and family relationships with networks and larger systems.

Visit the FPI Website for resource links, information and online discussions: www.family process.org

WILEY-
BLACKWELL

For more information and to subscribe online visit
www.blackwellpublishing.com/FAMP

NEW BOOKS FROM WILEY-BLACKWELL



Happiness
Unlocking the Mysteries of Psychological Wealth
ED DEINER and
ROBERT BISWAS-DIENER
August 2008 • HB
978-1-4051-4661-6

Women at the Top
Powerful Leaders Tell Us How to Combine Work and Family
DIANE F. HALPERN and
FANNY M. CHEUNG
August 2008 • HB
978-1-4051-7105-2

Raising Kids in the 21st Century
The Science of Psychological Health for Children
SHARON K. HALL
August 2008 • HB
978-1-4051-5805-3

Changing Realities of Work and Family
Edited by
AMY MARCUS-NEWHALL,
DIANE F. HALPERN, and
SHERYLLE J. TAN
2008 • PB
978-1-4051-6346-0

A History of Psychology
Original Sources and Contemporary Research
Third Edition
LUDY T. BENJAMIN, JR.
August 2008 • PB
978-1-4051-7710-8

Effective Writing in Psychology
Papers, Posters, and Presentations
BERNARD BEINS and
AGATHA BEINS
2008 • PB
978-1-4051-5879-4

Cyber Bullying
Bullying in the Digital Age
ROBIN M. KOWALSKI,
SUSAN P. LIMBER and
PATRICIA W. AGATSTON
2008 • PB
978-1-4051-5992-0

Children and Bullying
How Parents and Educators Can Reduce Bullying at School
KEN RIGBY
2008 • PB
978-1-4051-6254-8

Research Methods for Social Psychology
DANA S. DUNN
August 2008 • HB
978-1-4051-4980-8

Beyond Common Sense
Psychological Science in the Courtroom
Edited by
EUGENE BORGIDA and
SUSAN T. FISKE
2008 • PB
978-1-4051-4574-9

Forensic Psychology
Research, Clinical Practice, and Applications
MATTHEW T. HUSS
August 2008 • PB
978-1-4051-5138-2

Hypnosis
A Brief History
JUDITH PINTAR and
STEVEN J. LYNN
September 2008 • PB
978-1-4051-3452-1

Optimizing Teaching and Learning:
Pedagogical Research in Practice
REGAN A. R. GURUNG
and BETH M. SCHWARTZ
August 2008 • HB
978-1-4051-6179-4

The Science of Compassionate Love
BEVERLEY FEHR,
SUSAN SPRECHER, and
LYNN G. UNDERWOOD
August 2008 • PB
978-1-4051-5394-2

NEW SPSSI SERIES
Social Issues and Interventions

Explaining the Breakdown of Ethnic Relations
Why Neighbors Kill
Edited by
VICTORIA M. ESSES and
RICHARD A. VERNON
2008 • PB
978-1-4051-7058-1

Improving Intergroup Relations
Building on the Legacy of Thomas F. Pettigrew
Edited by
ULRICH WAGNER,
LINDA R. TROPP, GILLIAN
FINCHILESCU, and
COLIN TREDOUX
2008 • PB
978-1-4051-6971-4

Regulating Emotions
Culture, Social Necessity, and Biological Inheritance
MARIE VANDEKERCKHOVE,
CHRISTIAN VON
SCHEVE, SVEN ISMER,
SUSANNE JUNG, and
STEFANIE KRONAST
2008 • HB
978-1-4051-5863-3

NEW SERIES
Teaching Psychological Science

A Guide to Teaching Research Methods in Psychology
BRYAN K. SAVILLE
2008 • PB
978-1-4051-5481-9

A Guide to Teaching Introductory Psychology
SANDRA GOSS LUCAS
2008 • PB
978-1-4051-5151-1

A Guide to Teaching Developmental Psychology
ELIZABETH BRESTAN
KNIGHT and EMBER L. LEE
2008 • PB
978-1-4051-5781-0

A Guide to Teaching Statistics
Innovations and Best Practices
MICHAEL R. HULSIZER
and LINDA M. WOOLF
September 2008 • PB
978-1-4051-5574-8

WILEY-BLACKWELL

www.blackwellpublishing.com
Books Telephone: 800-216-2522 • Fax: 802-864-7626 **Journals** Telephone: 800-835-6770 • Fax: 781-388-8232

172015

POSTAL SERVICE® **(All Periodicals Publications Except Requester Publications)**

1. Publication Title
Child Development

2. Publication Number
8 8 9 _ 0 2 0

3. Filing Date
10/1/08

4. Issue Frequency
Bi-monthly

5. Number of Issues Published Annually
6

6. Annual Subscription Price
$585.00

7. Complete Mailing Address of Known Office of Publication *(Not printer) (Street, city, county, state, and ZIP+4®)*
Wiley Subscription Services, Inc., 111 River Street, Hoboken, NJ 07030

Contact Person
E. Schmidichen

Telephone *(Include area code)*
(201) 748-6346

8. Complete Mailing Address of Headquarters or General Business Office of Publisher *(Not printer)*
Wiley Subscription Services, Inc., 111 River Street, Hoboken, NJ 07030

9. Full Names and Complete Mailing Addresses of Publisher, Editor, and Managing Editor *(Do not leave blank)*

Publisher *(Name and complete mailing address)*
Wiley Subscription Services, Inc., 111 River Street, Hoboken, NJ 07030

Editor *(Name and complete mailing address)*
Lynn S. Liben, Pennsylvania State University, 450 Moore Bldg, University Park, PA 16802

Managing Editor *(Name and complete mailing address)*
Detra Davis, 2950 S. State Street, Suite 401, Ann Arbor, MI 48104

10. Owner *(Do not leave blank. If the publication is owned by a corporation, give the name and address of the corporation immediately followed by the names and addresses of all stockholders owning or holding 1 percent or more of the total amount of stock. If not owned by a corporation, give the names and addresses of the individual owners. If owned by a partnership or other unincorporated firm, give its name and address as well as those of each individual owner. If the publication is published by a nonprofit organization, give its name and address.)*

Full Name	**Complete Mailing Address**
Society for Research in Child Development	University of Michigan
	2950 S. State Street, Suite 401
	Ann Arbor, MI 48104

11. Known Bondholders, Mortgagees, and Other Security Holders Owning or Holding 1 Percent or More of Total Amount of Bonds, Mortgages, or Other Securities. If none, check box → ☑ None

Full Name	**Complete Mailing Address**

12. Tax Status *(For completion by nonprofit organizations authorized to mail at nonprofit rates) (Check one)*
The purpose, function, and nonprofit status of this organization and the exempt status for federal income tax purposes:
☐ Has Not Changed During Preceding 12 Months
☐ Has Changed During Preceding 12 Months *(Publisher must submit explanation of change with this statement)*

PS Form **3526**, September 2006 *(Page 1 of 3 (Instructions Page 3))* PSN 7530-01-000-9931 **PRIVACY NOTICE**: See our privacy policy on *www.usps.com*

13. Publication Title
Child Development

14. Issue Date for Circulation Data
September/October 2008

15. **Extent and Nature of Circulation**			**Average No. Copies Each Issue During Preceding 12 Months**	**No. Copies of Single Issue Published Nearest to Filing Date**
a. Total Number of Copies *(Net press run)*			8698	8670
b. Paid Circulation *(By Mail and Outside the Mail)*	(1)	Mailed Outside-County Paid Subscriptions Stated on PS Form 3541 *(Include paid distribution above nominal rate, advertiser's proof copies, and exchange copies)*	5761	6346
	(2)	Mailed In-County Paid Subscriptions Stated on PS Form 3541 *(Include paid distribution above nominal rate, advertiser's proof copies, and exchange copies)*	0	0
	(3)	Paid Distribution Outside the Mails Including Sales Through Dealers and Carriers, Street Vendors, Counter Sales, and Other Paid Distribution Outside USPS®	0	0
	(4)	Paid Distribution by Other Classes of Mail Through the USPS (e.g. First-Class Mail®)	0	0
c. Total Paid Distribution *(Sum of 15b (1), (2),(3), and (4))*			5761	6346
d. Free or Nominal Rate Distribution *(By Mail and Outside the Mail)*	(1)	Free or Nominal Rate Outside-County Copies included on PS Form 3541	99	98
	(2)	Free or Nominal Rate In-County Copies Included on PS Form 3541	0	0
	(3)	Free or Nominal Rate Copies Mailed at Other Classes Through the USPS (e.g. First-Class Mail)	0	0
	(4)	Free or Nominal Rate Distribution Outside the Mail *(Carriers or other means)*	0	0
e. Total Free or Nominal Rate Distribution *(Sum of 15d (1), (2), (3) and (4)*			99	98
f. Total Distribution *(Sum of 15c and 15e)* ▶			5860	6444
g. Copies not Distributed *(See Instructions to Publishers #4 (page #3))* ▶			2838	2226
h. Total *(Sum of 15f and g)* ▶			8698	8670
i. Percent Paid *(15c divided by 15f times 100)* ▶			98.31	98.48

16 Publication of Statement of Ownership
☑ If the publication is a general publication, publication of this statement is required. Will be printed in the December 2008 issue of this publication.
☐ Publication not required.

17. Signature and Title of Editor, Publisher, Business Manager, or Owner
Elizabeth Konkle, Associate Financial Manager

Date
10/1/08

I certify that all information furnished on this form is true and complete. I understand that anyone who furnishes false or misleading information on this form or who omits material or information requested on the form may be subject to criminal sanctions (including fines and imprisonment) and/or civil sanctions (including civil penalties).

Notice to Contributors

Child Development publishes empirical, theoretical, review, applied, and policy articles reporting research on child development. Published by the interdisciplinary Society for Research in Child Development (SRCD), the journal welcomes relevant submissions from all disciplines. Further information is available at http://www.srcd.org/cd.html.

Types of Articles

Child Development considers manuscripts in formats described below. Inquiries concerning alternative formats should be addressed to the Editor prior to submission. All submissions are expected to be no more than 40 manuscript pages, including tables, references, and figures (but excluding appendices). Authors should provide a justification if the submission is substantially longer. Unless the editor finds that justification compelling, the submission will be returned to the author for shortening prior to editorial review.

Empirical articles comprise the major portion of the journal. To be accepted, empirical articles must be judged as being high in scientific quality, contributing to the empirical base of child development, and having important theoretical, practical, or interdisciplinary implications. Reports of multiple studies, methods, or settings are encouraged, but single-study reports are also considered. Empirical articles will thus vary considerably in length, but should be no longer than 40 manuscript pages; text and graphics should be as concise as material permits. All modes of empirical research are welcome.

Empirical reports are reserved for short cutting-edge empirical papers that are no longer than 4,000 words in length (including text, tables, footnotes, appendices), which advance research and knowledge in an area through noteworthy findings and/or new methods. For manuscripts that require longer descriptions of methods and results, authors should use the Empirical article format.

Reviews focus on past empirical and/or on conceptual and theoretical work. They are expected to synthesize or evaluate a topic or issue relevant to child development, should appeal to a broad audience, and may be followed by a small number of solicited commentaries.

Essays describe original concepts, methods, trends, applications, and theories; these may also be accompanied by solicited commentaries.

Child Development and … are articles that provide readers with tutorials about some new concept or academic specialty pertinent to research in child development. These papers should review the major definitions, methods, and findings of the concept or specialty and discuss past or potential links to child development.

From another perspective is a format in which papers on a focal topic, written by different authors, are published simultaneously. Papers represent diverse perspectives (e.g., authors whose work represents different populations; different disciplines; different theories, methods, or analytic tools). In some cases, calls for submissions on particular topics will be disseminated through SRCD (via e-mail or SRCD publications), and submissions will undergo normal editorial review. In some cases, a submitted manuscript (e.g., an empirical article) may be selected as a lead article for this format, with invited commentaries providing additional perspectives. The editors also welcome suggestions from readers for topics for this format.

Manuscript Submission

Child Development invites for consideration manuscripts that are neither identical to nor substantially similar to work published or under review elsewhere. In the submission cover letter, please provide details about other published or submitted papers having substantial overlap (including data sets) with the new CD submission to enable editors to judge whether the new submission is sufficiently distinct from other work to warrant consideration. Please note if the paper is posted on a website, see http://www. srcd.org/webposting.html. Editors retain the right to reject manuscripts that do not meet established ethical standards for research or dissemination.

Manuscripts should be submitted electronically to the Child Development Online Submission Site at www.srcd.org/CDsubmit/ as a Word or WordPerfect file. Please also submit a cover letter that contains the name(s) of the author(s) and affiliation(s), and the street address, telephone, fax, and electronic mail address of the corresponding author. A corresponding author's submission to *Child Development* implies that all coauthors have agreed to the content and form of the manuscript and that the ethical standards of SRCD have been followed (see the *Child Development* website or pp. 283–284 of the 2000 SRCD Directory). Any financial interest or conflict of interest must be explained to the Editor in the cover letter. The corresponding author is responsible for informing all coauthors, in a timely manner, of manuscript submission, editorial decisions, reviews, and revisions. Please note that if your submission (or its revision) is accepted for publication, each author must submit a signed Exclusive License Form(http://www.srcd.org/Documents/Publications/CD/cdcopyright.pdf) before the manuscript can enter production. Planning ahead to meet this requirement is especially important for manuscripts with multiple authors.

The manuscript file should be formatted with double-spaced, 12-point type, and should include a single paragraph abstract of 100–120 words. Please follow all guidelines on format, style, and ethics provided in the Publication Manual (5th ed.) of the American Psychological Association. Figures included with initial submissions will not be returned. Therefore, please submit only electronic files or copies of figures. Authors should keep a copy of all correspondence, files, and figures to guard against loss.

Manuscript Review

If you have any questions about your submission, please inquire at cdev@srcd.org or call (734) 926-0615. Each manuscript is handled by the Editor or an Associate Editor who consults with one or more Consulting Editors and/or ad hoc reviewers who have relevant expertise. To ensure blind review, cover sheets are removed before review; authors should avoid including any other information about identity or affiliation in submissions. Copies of the submission and associated correspondence are retained in the SRCD archives. For accepted manuscripts, authors are required to prepare a 300–500 layperson's summary for public dissemination purposes. Details are provided to authors as part of final processing.

There is no charge for publication in *Child Development* unless tabular or graphic materials exceed 10% of the total number of pages. Charges are also levied for changes in proofs other than correction of printer's errors. Any inquiries relating to charges or business matters (including reprint orders) should be addressed to Wiley-Blackwell, *Child Development*, Production Coordinator, 350 Main Street, Malden, MA 02148, (781) 388-8200.

Inquiries and suggestions regarding editorial policy may be addressed to: Dr. Jeffrey J. Lockman, Incoming Editor, Department of Psychology, Tulane University, 2007 Percival Stern Hall, New Orleans, LA 70118-5698, lockman@tulane.edu. *Please do not send electronic or paper manuscripts to the above address.* Doing so will delay the processing of your manuscript.

Note to NIH Grantees

Pursuant to NIH mandate, Society through Wiley-Blackwell will post the accepted version of Contributions authored by NIH grantholders to PubMed Central upon acceptance. This accepted version will be made publicly available 12 months after publication. For further information, see www.wiley.com/go/nihmandate.

Monographs of the Society for Research in Child Development

Recent

Understanding Mother-Adolescent Conflict Discussions: Concurrent and Across-Time Prediction From Youths' Dispositions and Parenting—*Nancy Eisenberg, Claire Hofer, Tracy L. Spinrad, Elizabeth T. Gershoff, Carlos Valiente, Sandra Losoya, Qing Zhou, Amanda Cumberland, Jeffrey Liew, Mark Reiser, and Elizabeth Maxon* (Serial No. 290, 2008)

Developing Object Concepts in Infancy: An Associative Learning Perspective—*David H. Rakison and Gary Lupyan* (Serial No. 289, 2008).

The Genetic and Environmental Origins of Learning Abilities and Disabilities in the Early School Years—*Yulia Kovas, Claire M. A. Haworth, Philip S. Dale, and Robert Plomin* (Serial No. 288, 2007).

The Preservation of Two Infant Temperaments Into Adolescence—*Jerome Kagan, Nancy Snidman, Vali Kahn, and Sara Towsley* (Serial No. 287, 2007).

Children's Questions: A Mechanism for Cognitive Development—*Michelle M. Chouinard* (Serial No. 286, 2007).

Best Practices in Quantitative Methods for Developmentalists—*Kathleen McCartney, Margaret R. Burchinal, and Kristen L. Bub* (Serial No. 285, 2006).

Foundations of Self-Awareness: An Exploration Through Autism—*R. Peter Hobson, Gayathri Chidambi, Anthony Lee, and Jessica Meyer* (Serial No. 284, 2006).

Our Children Too: A History of the Black Caucus of the Society for Research in Child Development, 1973–1997—*Diana T. Slaughter-Defoe, Aline M. Garrett, and Algea O. Harrison-Hale* (Serial No. 283, 2006).

Parental Support, Psychological Control, and Behavioral Control: Assessing Relevance Across Time, Culture, and Method—*Brian K. Barber, Heidi E. Stolz, and Joseph A. Olsen* (Serial No. 282, 2005).

Being Hurt and Hurting Others: Children's Narrative Accounts and Moral Judgments of Their Own Interpersonal Conflicts—*Cecilia Wainryb, Beverly A. Brenl, and Sonia Matwin* (Serial No. 281, 2005).

Childhood Sexual Assault Victims: Long-Term Outcomes After Testifying in Criminal Court—*Jodi A. Quas, Gail S. Goodman, Simona Ghetti, Kristen Weede Alexander, Robin Edelstein, Allison D. Redlich, Ingrid M. Cordon, and David P. H. Jones* (Serial No. 280, 2005).

The Emergence of Social Cognition in Three Young Chimpanzees—*Michael Tomasello and Malinda Carpenter* (Serial No. 279, 2005).

Trajectories of Physical Aggression From Toddlerhood to Middle Childhood: Predictors, Correlates, and Outcomes—*NICHD Early Child Care Research Network* (Serial No. 278, 2004).

Constraints on Conceptual Development: A Case Study of the Acquisition of Folkbiological and Folksociological Knowledge in Madagascar—*Rita Astuti, Gregg E. A. Solomon, and Susan Carey* (Serial No. 277, 2004).

Origins and Early Development of Human Body Knowledge—*Virginia Slaughter, Michelle Heron, Linda Jenkins, and Elizabeth Tilse* (Serial No. 276, 2004).

Mother – Child Conversations About Gender: Understanding the Acquisition of Essentialist Beliefs—*Susan A. Gelman, Marianne G. Taylor, and Simone P. Nguyen* (Serial No. 275, 2004).

The Development of Executive Function in Early Childhood—*Philip David Zelazo, Ulrich Müller, Douglas Frye, and Stuart Marcovitch* (Serial No. 274, 2003).

Personal Persistence, Identity Development, and Suicide: A Study of Native and Non-native North American Adolescents—*Michael J. Chandler, Christopher E. Lalonde, Bryan W. Sokol, and Darcy Hallett* (Serial No. 273, 2003).

Personality and Development in Childhood: A Person-Centered Approach—*Daniel Hart, Robert Atkins, and Suzanne Fegley* (Serial No. 272, 2003).

How Children and Adolescents Evaluate Gender and Racial Exclusion—*Melanie Killen, Jennie Lee-Kim, Heidi McGlothlin, and Charles Stangor* (Serial No. 271, 2002).

Child Emotional Security and Interparental Conflict—*Patrick T. Davies, Gordon T. Harold, Marcie C. Goeke-Morey, and E. Mark Cummings* (Serial No. 270, 2002).

The Developmental Course of Gender Differentiation: Conceptualizing, Measuring, and Evaluating Constructs and Pathways—*Lynn S. Liben and Rebecca S. Bigler* (Serial No. 269, 2002).

The Development of Mental Processing: Efficiency, Working Memory, and Thinking—*Andreas Demetriou, Constantinos Christou, George Spanoudis, and Maria Platsidou* (Serial No. 268, 2002).

Order from Wiley-Blackwell at subscrip@bos.blackwellpublishing.com

This journal is available online at Wiley InterScience. Visit www3.interscience.wiley.com to search the articles and register for table of contents e-mail alerts.

MARYGROVE
LIBRARY
DETROIT, MI.